RESEARCH

Comparison and analysis of deep learning models for discriminating longitudinal and oblique vaginal septa based on ultrasound imaging

Xiangyu Wang[1†], Liang Wang[2†], Xin Hou[1], Jingfang Li[1], Jin Li[2*†] and Xiangyi Ma[1*†]

Abstract

Background The longitudinal vaginal septum and oblique vaginal septum are female müllerian duct anomalies that are relatively less diagnosed but severely fertility-threatening in clinical practice. Ultrasound imaging is commonly used to examine the two vaginal malformations, but in fact it's difficult to make an accurate differential diagnosis. This study is intended to assess the performance of multiple deep learning models based on ultrasonographic images for distinguishing longitudinal vaginal septum and oblique vaginal septum.

Methods The cases and ultrasound images of longitudinal vaginal septum and oblique vaginal septum were collected. Two convolutional neural network (CNN)-based models (ResNet50 and ConvNeXt-B) and one base resolution variant of vision transformer (ViT)-based neural network (ViT-B/16) were selected to construct ultrasonographic classification models. The receiver operating curve analysis and four indicators including accuracy, sensitivity, specificity and area under the curve (AUC) were used to compare the diagnostic performance of deep learning models.

Results A total of 70 cases with 426 ultrasound images were included for deep learning models construction using 5-fold cross-validation. Convolutional neural network-based models (ResNet50 and ConvNeXt-B) presented significantly better case-level discriminative efficacy with accuracy of 0.842 (variance, 0.004, 95%CI, [0.639–0.997]) and 0.897 (variance, 0.004, [95%CI, 0.734-1.000]), specificity of 0.709 (variance, 0.041, [95%CI, 0.505–0.905]) and 0.811 (variance, 0.017, [95%CI, 0.622–0.979]), and AUC of 0.842 (variance, 0.004, [95%CI, 0.639–0.997]) and 0.897 (variance, 0.004, [95%CI, 0.734-1.000]) than transformer-based model (ViT-B/16) with its accuracy of 0.668 (variance, 0.014, [95%CI, 0.407–0.920]), specificity of 0.572 (variance, 0.024, [95%CI, 0.304–0.831]) and AUC of 0.681 (variance, 0.030, [95%CI, 0.434–0.908]). There was no significance of AUC between ConvNeXt-B and ResNet50 (*P* = 0.841).

---

†Xiangyu Wang and Liang Wang contributed equally to this work and share the first authorship.

†Jin Li and Xiangyi Ma contributed equally to this work and share the corresponding authorship.

*Correspondence:
Jin Li
lijin@tjh.tjmu.edu.cn
Xiangyi Ma
xyma@tjh.tjmu.edu.cn
Full list of author information is available at the end of the article

Wang *et al. BMC Medical Imaging*   (2024) 24:347

Page 2 of 12

Conclusions Convolutional neural network-based model (ConvNeXt-B) shows promising capability of discriminating longitudinal and oblique vaginal septa ultrasound images and is expected to be introduced to clinical ultrasonographic diagnostic system.

Keywords Longitudinal vaginal septum, Oblique vaginal septum, Ultrasound imaging, Deep learning model, Classification

Introduction

Müllerian duct anomalies (MDAs) are a collection of female congenital malformations probably involving uterus, cervix and vagina, with an incidence of about 5%, which have negative impacts on women's physiology and psychology [1]. It's reported that the prevalence of MDAs varies from 6 to 15% in women disturbed by fertility concerns and recurrent miscarriages, rising to 25% in women suffering both infertility and recurrent pregnancy loss [2, 3]. Oblique vaginal septum belongs to an important subgroup of vaginal malformations of MDAs. According to the American Society for Reproductive Medicine (ASRM) classification of MDAs, it is defined as oblique vaginal septum when asymmetric septum and obstructed hemivagina occur, frequently co-existent with didelphic uterus/septate uterus and ipsilateral renal agenesis [4–6]. The clinical presentations of oblique vaginal septum are quite diverse and severe, such as dysmenorrhea, vaginal/pelvic mass and cyclic lower abdomen pain, etc [5, 7, 8]. Relatively, longitudinal vaginal septum is another main group of vaginal malformations, with a septum of variable length, and patients with longitudinal vaginal septum are commonly asymptomatic or present with dyspareunia [9]. Hence, considering the reproductive threatening and pain bearing, precisely identifying the longitudinal and oblique vaginal septa is of much importance.

In clinical practice, three-dimensional ultrasonography (3D-US) tends to be the diagnostic imaging of first-choice for vaginal anomalies, with comparable accuracy, greater accessibility, better operation convenience and lower economic cost [10–12]. However, due to the rarity of incidence, complexity of anatomical structures, and confusion of clinical manifestations of oblique vaginal septum and longitudinal vaginal septum, the inexperienced sonographers are likely to make a missed diagnosis and misdiagnosis, leading to the delay in treatment and increased risk of some severe complications including endometriosis, pelvic adhesions and infertility [13–15]. So, there is an urgent need to optimize the ultrasonography based diagnostic modality of longitudinal vaginal septum and oblique vaginal septum, to aid the sonographers make precise diagnosis and the clinicians make appropriate interventional decision-making.

In the past decade, deep learning (DL) technology has played an increasingly crucial role in medical images analysis [16, 17]. CNN as a powerful DL algorithm, has shown outstanding talent to acquire the representative feature of images automatically, making DL-assisted diagnostic decision-making viable. It's reported that about 80% studies of medical images analysis involved the CNN [18, 19]. In obstetrical and gynecological congenital malformation field, CNN-based models have been widely applied in fetal ultrasound imaging classification. Xie et al. [19] utilized CNN-based DL algorithms to identify abnormal fetal brain ultrasound images. Li et al. [20] proposed a novel CNN-based model to recognize fetal lip abnormalities in ultrasound images accurately. Transformer is designed to conduct nature language processing or computer vision, which has attracted attention in medical image analysis as well [21]. Recently, transformer-based models have been introduced to improve the recognizing of fetal congenital heart disease [22].

However, the discrimination of longitudinal and oblique vaginal septa in ultrasound images using DL techniques has not been reported. In this study, two CNN [17]-based models (ResNet50 and ConvNeXt-B) and one transformer-based model (ViT-B/16) will be used for the classification of ultrasound images of longitudinal vaginal septum and oblique vaginal septum. Among them, ResNet50 was the most commonly used CNN-based DL model [23]. ViT-B/16 was a transformer-based DL model, and ConvNeXt-B was created by introducing part of the idea of transformer into CNN through depthwise convolution to achieve a similar effect [24]. With the assistance of deep learning models, we hope to achieve accurate ultrasonographic diagnosis of longitudinal and oblique vaginal septa and to improve treatment decision-making.

Materials and methods

Data collection

This is a retrospectively analytic study. We collected accessible cases with a diagnosis of longitudinal vaginal septum or oblique vaginal septum in the Department of Obstetrics and Gynecology, Tongji Hospital, Wuhan between January 2013 and January 2023. The ultrasonographic images were digitally stored in the computer center of our hospital. Patients with ultrasonographic images for analysis must simultaneously meet the following criteria for inclusion: (1) patients surgically diagnosed with a longitudinal or oblique vaginal septum; (2) patients with gynecological ultrasound imaging record indicating longitudinal/oblique vaginal septum in our hospital; (3) only preoperative ultrasound imaging

Wang *et al. BMC Medical Imaging*   (2024) 24:347

Page 3 of 12

of longitudinal/oblique vaginal septum will be adopted; (4) only B-mode images will be used. The exclusion criteria are as follows: (1) patients without longitudinal/oblique vaginal septum; (2) patients without ultrasound imaging records; (3) postoperative ultrasound imaging; (4) ultrasound imaging during pregnancy; (5) ultrasound images that are not in B-mode or with low quality. The flowchart of data collection and inclusion is shown in Fig. 1.

Ultrasonographic image segmentation

The original ultrasonographic images contain noisy information at the edge that may result in unsatisfactory performance of the DL models. Besides, previous studies have revealed that some key features like hematocolpos were beneficial to define oblique vaginal septum [8]. The longitudinal and oblique vaginal septa were frequently accompanied by uterus anomalies, such as septate uterus and uterus didelphys [25, 26]. Therefore, the images were segmented by using ITK-SNAP 4.0 with a red minor rectangular or trapezoid mask to annotate the region of interest (ROI) containing uterus and vagina and simultaneously to exclude extra interfering information as much as possible (Fig. 2).
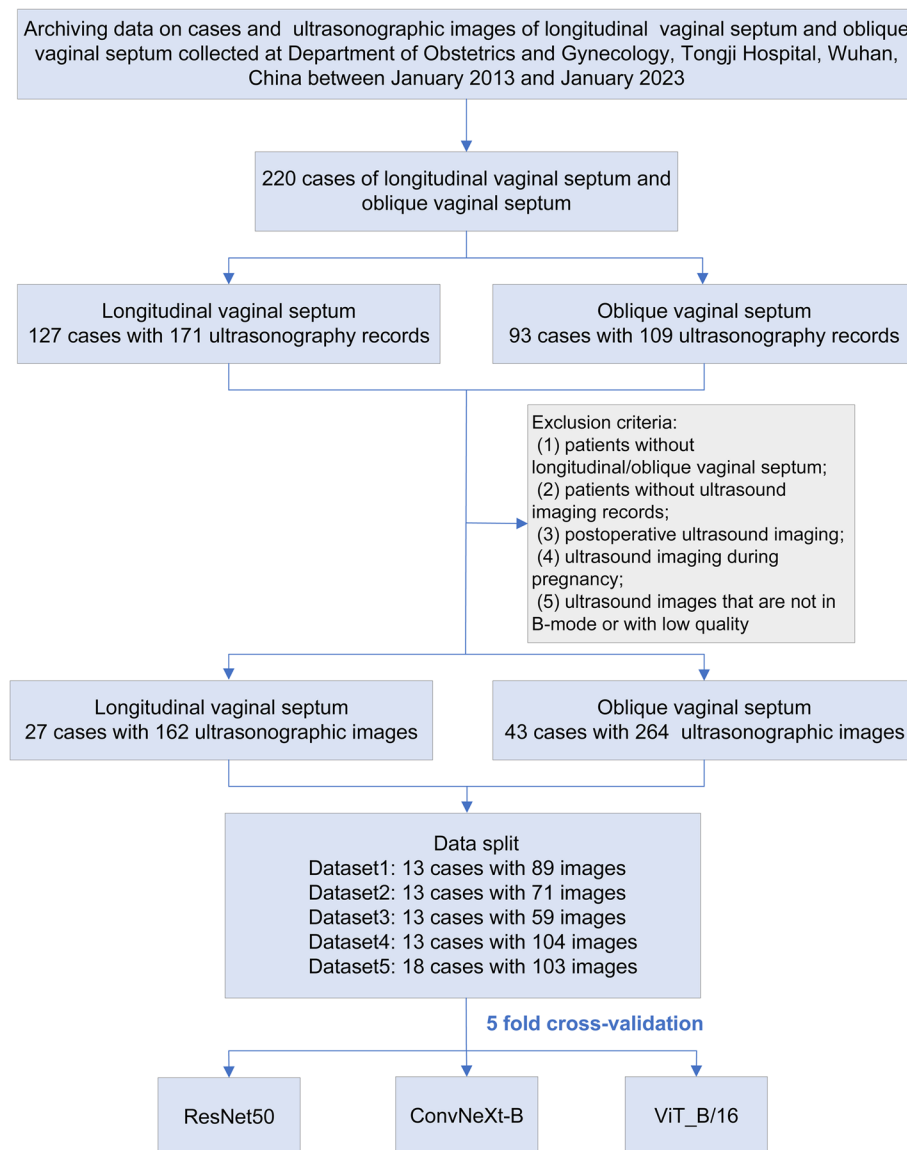
Fig. 1 Flowchart of data collection for multiple DL models construction. Case data and ultrasonographic images were collected after receiving the ethical approval in Tongji hospital, Wuhan. Strict exclusion criteria were established to enroll the data. 5- fold cross-validation was used to construct deep learning models. DL, deep learning

Wang *et al. BMC Medical Imaging*   (2024) 24:347

Page 4 of 12
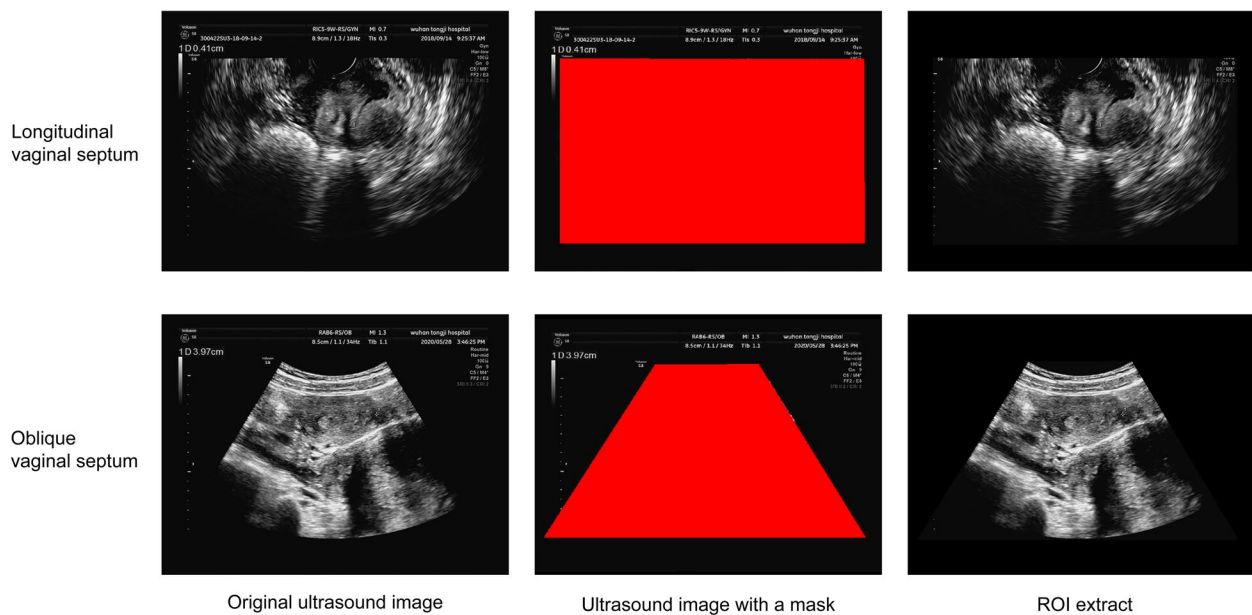
Fig. 2 Illustration of ultrasound images segmentation of longitudinal and oblique vaginal septa. Original ultrasound images of longitudinal and oblique vaginal septa were collected. Then red mask was used to mark the region of uterus and vagina, which was defined as the region of interest and taken as input information of the deep learning models. ROI, region of interest

Deep learning neural networks

Three DL models were selected for image classification, among which ResNet50 is the most commonly used CNN-based DL model [23]. ViT-B/16 is a transformer-based DL model, and by co-opting partial idea of transformer, ConvNeXt-B can achieve a similar effect through depthwise convolution [24, 27].

ResNet50 is a representative network of the ResNets. As shown in Fig. 3a, ResNet50 consists of six layers, including input layer, convolution layer, pooling layer, residual block, global average pooling layer, and output layer. The core idea is to introduce short-circuit connections to convert learning objectives into residuals, thereby reducing the impact on model performance when increasing network depth.

ResNet50 is one of the most widely used deep learning models in image classification, of which the advantage is that it solves the gradient vanishing problem in deep networks through residual connections, allowing deeper networks to be trained, and showing good performance on a variety of tasks and data sets. But due to more reliance on capturing local features, it is weak in capturing global features of images.

ConvNeXt comes in multiple versions, with variations in the number of channels and blocks within each stage. ConvNeXt is composed of an image preprocessing layer, a processing layer, and a classification layer. Its core lies in the processing layer, which contains four stages. Each stage repeats several blocks, and the number of repetitions defines the depth of the model. The aim is to extract image features through a series of convolutional operations. In ConvNeXt-B, the input channel counts for the four stages respectively are 128, 256, 512 and 1024, and the number of block repetitions respectively are 3, 3, 27 and 3 (Fig. 3b).

Compared with ResNet50, the advantage of ConvNeXt-B is that it uses large convolution kernels to expand the receptive field, thereby capturing more contextual information. At the same time, through the introduction of deep separable convolution, the model maintains high performance while relatively reducing the consumption of computing resources. The disadvantage is that the use of large convolution kernels requires more memory usage.

ViT-B/16 is a variant of ViT-B, with a patch size of 16×16, and ViT-B is the base resolution variant in the ViT model. The core modules of the ViT model consist of three parts: the embedding layer, the transformer encoder, and the multilayer perceptron (MLP) head. In the embedding layer, the input image is divided into a series of image patches, and each image patch is represented as a sequence of vectors after undergoing a linear mapping. In the transformer encoder, ViT utilizes multiple transformer encoder layers to process the input sequence of image patches. Each transformer encoder layer comprises a self-attention mechanism and a feed-forward network. The self-attention mechanism captures the relationships between image patches

Wang *et al. BMC Medical Imaging*   (2024) 24:347

Page 5 of 12
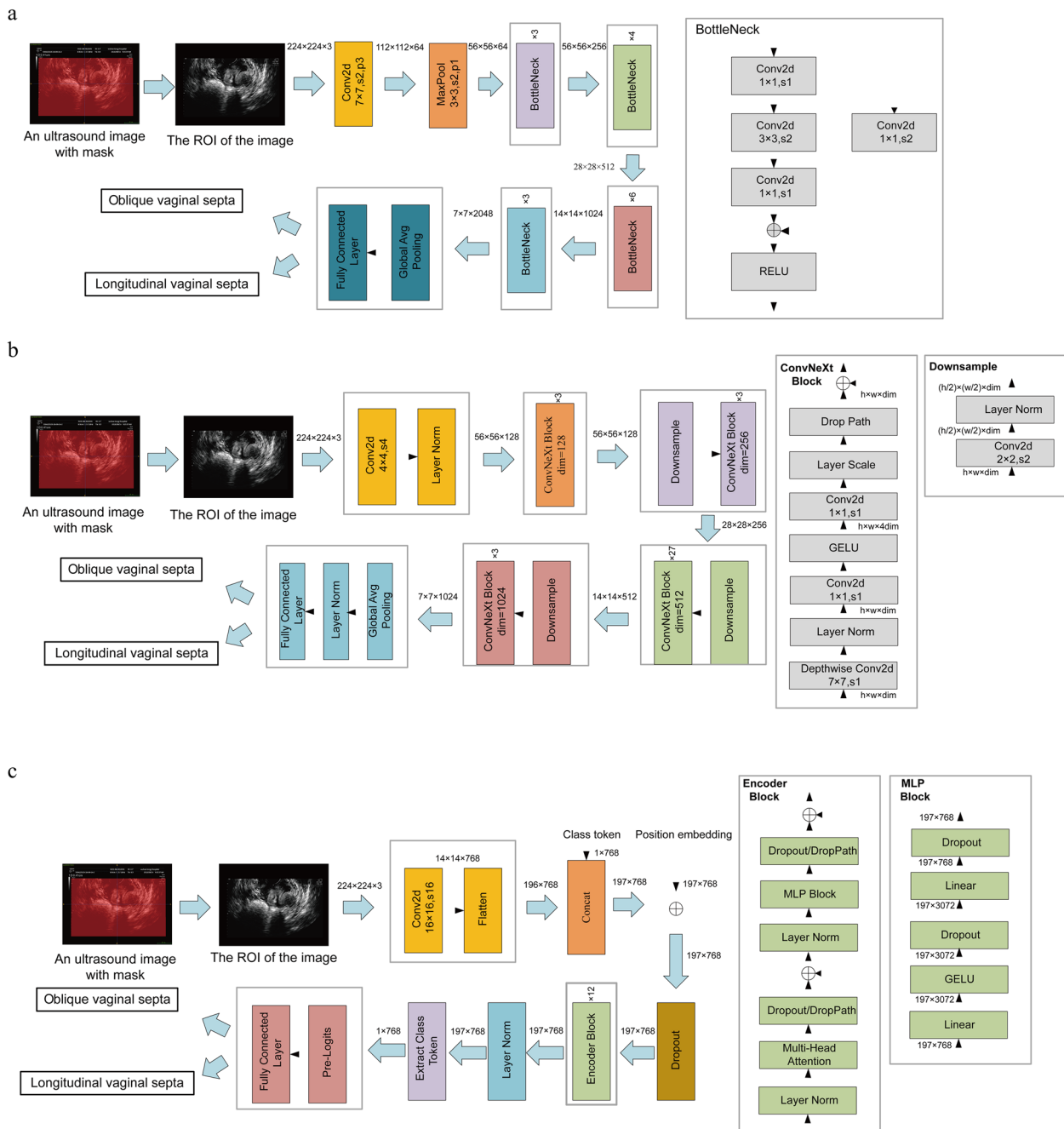
Fig. 3 DL model neural network architectures for discriminating longitudinal and oblique vaginal septa. **a**. The neural network architecture of ResNet50; **b**. The neural network architecture of ConvNeXt-B; **c**. The neural network architecture of ViT-B/16. DL, deep learning

and generates context-aware feature representations for them. In the MLP head, the output of the ViT model is used for classification prediction through an additional linear layer. Typically, a global average pooling operation is applied to the output of the last transformer encoder layer, aggregating the features of the image patch sequence into a global feature vector, which is then classified using a linear layer (Fig. 3c).

The advantage of ViT-B/16 is that the transformer architecture used can capture global dependencies in images through the self-attention mechanism. Since it does not rely on convolutional layers, the model can be easily extended to different image sizes and resolutions; its disadvantage is requiring more data to train. After the ultrasound image is input into the model, the features extracted by each layer of the ViT-B/16 network

Wang *et al. BMC Medical Imaging*   (2024) 24:347

Page 6 of 12

are a mixture of local and global information, while the ResNet50 and ConvNeXt-B networks extract local features first and then refine global features using local features.

Model construction

We constructed DL models for the differential diagnosis of longitudinal vaginal septum and oblique vaginal septum using 5-fold cross-validation. The size of each input ultrasound image was uniformly scaled to 224×224 and normalized for each channel. In 5-fold cross-validation, all patients were divided into five parts, with ultrasound images of the same individual in the same part, and the proportion of individuals with longitudinal vaginal septum and oblique vaginal septum kept constant in each part. During the training of each fold, one part was selected as the validation set, and the other four parts were used as the training set. Besides, the hyperparameters including learning rate for model training were set as shown in Table S1. The optimizer used is Adam, which calculates the first and second moments of the gradient to adjust the learning rate for each parameter of models. Betas are the exponential decay factors in Adam that can be used to control the adjustment of learning rate. Eps is a constant added to the denominator to prevent division by zero errors, and weight-decay is the L2 regularization term, which is used to control the magnitude of parameters to prevent overfitting. The hyperparameter epoch was adjusted to a suitable value so that the loss of the training set reached plateau. During the training process, the optimal model parameters were selected according to the highest accuracy achieved on the training set.

In addition, we further considered combining the three well-trained DL models, using the outputs of each model as inputs for the XGBoost classification algorithm to construct a combined classification model. By utilizing the Optuna library in Python, we can search the XGBoost hyperparameter space based on the Bayesian optimization algorithm, and automatically obtain the best model hyperparameters. The parameter settings of Optuna are as shown in Table S2.

Model visualization

To visually compare the differences between different models, the Grad-CAM package in Python was used to show the local focused regions of each model.

Statistical analysis

Quantitative data were compared using *t* test or Mann-Whitney U while qualitative data were compared using chi-square test. Receiver operating characteristic (ROC) curves were plotted to evaluate and compare the performance of DL models. Four metrics including accuracy, sensitivity, specificity and AUC were selected to quantify the performance of the models. 95% confidence intervals (95%CIs) of the numeric variables were calculated. The average value of these metrics was taken as the final performance estimated value. The difference was considered statistically significant with $P < 0.05$. The data was analyzed using Python 3.10.8.

Results

Basic clinical characteristics

A total of 27 longitudinal vaginal septum patients (mean age ± SD, 29 ± 6 years), with 164 ultrasonographic images and 43 oblique vaginal septum patients (mean age ± SD, 21 ± 8 years) with 262 ultrasonographic images were finally included for constructing deep learning models (Table S3).

Evaluation and comparison of image-based models

We constructed three DL models based on 70 cases with 426 ultrasonographic images. In the validation set, ConvNeXt-B had the best AUCof 0.897 (variance, 0.005, [95%CI, 0.843–0.949]) with highest accuracy of 0.834 (variance, 0.008, 95%CI, 0.762–0.905) and highest specificity of 0.845 (variance, 0.010, [95%CI, 0.778–0.911]). ResNet50 had an AUC of 0.876 (variance, 0.004, [95%CI, 0.813–0.936]) secondary only to ConvNeXt-B, with the best sensitivity of 0.829 (variance, 0.019, [95%CI, 0.769–0.890]). Relatively, the AUC of ViT-B/16 was only 0.664 (variance, 0.011, [95%CI, 0.567–0.761]), junior to both ResNet50 and ConvNeXt-B, with the worst performance, of which the accuracy, sensitivity and specificity were all lower than 0.800 (Table 1).

Further, multiple comparisons of diagnostic performance on validation set between any two of the three models were conducted. The accuracy and AUC of ConvNextB in image-level significantly outperformed ViT-B/16 with *P* value of 0.046 and 0.028, respectively. But there was no significant difference of the four metrics between ResNet50 and ViT-B/16 or ResNet50 and ConvNextB (Table 2).

We also considered combining the three DL models. The combined DL model with AUC of 0.891 (variance, 0.005, [95%CI, 0.837–0.944]) did not outperformed ConvNeXt-B at image-level (Table. S4).

Evaluation and comparison of the discriminative performance of case-based models

Besides, we constructed the DL models in case-based manner using 5-fold cross-validation. The ROC on validation set in each fold of the 5-fold cross-validation is shown in Fig. 4a. It's obvious that the AUCs of the models had variations among different data divisions. Therefore, for each performance metric of the model, the average of the metric

Wang *et al. BMC Medical Imaging*   (2024) 24:347

Page 7 of 12

Table 1 Discriminative performance of image-based DL models according to 5-fold cross-validation

Model	Dataset	Accuracy [Variance, (95%CI)]	Sensitivity [Variance, (95%CI)]	Specificity [Variance, (95%CI)]	AUC [Variance, (95%CI)]
ResNet50	Training	0.999 (0.000, [0.997–1.000])	0.998 (0.000, [0.996–1.000])	1.000 (0.000, [1.000–1.000])	1.000 (0.000, [1.000–1.000])
ResNet50	Validation	0.813 (0.004, [0.734–0.892])	**0.829 (0.019, [0.769–0.890])**	0.790 (0.007, [0.707–0.872])	0.876 (0.004, [0.813–0.936])
ConvNeXt-B	Training	1.000 (0.000, [1.000–1.000])	1.000 (0.000, [1.000–1.000])	1.000 (0.000, [1.000–1.000])	1.000 (0.000, [1.000–1.000])
ConvNeXt-B	Validation	**0.834 (0.008, [0.762–0.905])**	0.823 (0.010, [0.758–0.888])	**0.845 (0.010, [0.778–0.911])**	**0.897 (0.005, [0.843–0.949])**
ViT_B/16	Training	0.999 (0.000, [0.998–1.000])	0.999 (0.000, [0.998–1.000])	1.000 (0.000, [1.000–1.000])	0.999 (0.000, [0.998–1.000])
ViT_B/16	Validation	0.650 (0.006, [0.551–0.749])	0.706 (0.013, [0.614–0.797])	0.578 (0.006, [0.476–0.680])	0.664(0.011, [0.567–0.761])

Abbreviation: *DL* deep learning, *95%CI* 95% confidence interval, *AUC* area under the curve

Table 2 Comparison of the discriminative performance of image-based DL models on validation set

Model 1	Model 2	P_value (Accuracy)	P_value (Sensitivity)	P_value (Specificity)	P_value (AUC)
ResNet50	ViT_B/16	0.168	0.517	0.395	0.094
ConvNeXt-B	ViT_B/16	**0.046**	0.517	0.09	**0.028**
ResNet50	ConvNeXt-B	0.237	0.916	0.527	0.295

Abbreviation: *DL* deep learning, *AUC* area under the curve

values on validation set of each fold is considered as the final metric. Remarkably, ConvNeXt-B demonstrated the best AUC of 0.897 (variance, 0.004, [95%CI, 0.734–1.000]), with highest accuracy of 0.817 (variance, 0.004, [95%CI, 0.598–0.994]), sensitivity of 0.817 (variance, 0.016, [95%CI, 0.632–0.984]), and specificity of 0.811 (variance, 0.017, [95%CI, 0.622–0.979]). ResNet50 had a median AUC of 0.842 (variance, 0.004, 95%CI, 0.639–0.997) and relatively worse accuracy of 0.772 (variance, 0.005, 95%CI, 0.534–0.988), sensitivity of 0.808 (variance, 0.030, 95%CI, 0.631–0.960), and specificity of 0.709 (variance, 0.041, 95%CI, 0.505–0.905). Still, ViT-B/16 showed the worst performance with its all metrics less than 0.800 (Table 3).

We then compared four metrics of case-based DL models in validation set (Table 4). Except the sensitivity, the accuracy, specificity and AUC of ResNet50 significantly outperformed ViT-B/16 (*P* = 0.016, *P* = 0.032 and *P* < 0.01, respectively). Likewise, compared to ViT-B/16, ConvNeXt-B had significantly greater accuracy (*P* = 0.036), specificity (*P* < 0.01) and AUC (*P* < 0.01), but the sensitivity (*P* = 0.249). The four metrics of ConvNeXt-B and ResNet5 showed no significant differences (*P* > 0.05) (Table 4).

To figure out whether the combined DL model will behave better with case classification. We analyzed its case-level performance. Still, the combined DL model with specificity of 0.743 (variance, 0.022, [0.533–0.945]) and AUC of 0.891 (variance, 0.003, [95%CI, 0.736–1.000]), failed to identify the cases as well as ConvNeXt-B (Table. S4).

Model interpretation
Since the models had relatively better performance on the validation set in the third fold of the 5-fold cross-validation, we visualized the deep features based on the models trained in the third fold (Fig. 5).

Discussion
In this study, we totally introduced three DL deep learning algorithms to establish ultrasonographic diagnostic models to classify longitudinal vaginal septum and oblique vaginal septum. Importantly, we found that ConvNeXt-B showed the best capability to discriminate these two vaginal anomalies with the highest AUC of 0.897.

The longitudinal vaginal septum and oblique vaginal septum clinically belongs to two classifications of female vaginal anomalies [4]. Longitudinal vaginal septum may be asymptomatic in adolescence, thus avoiding medical examination, which will cause fertility problems and make the transvaginal delivery challenging in women of child-bearing age. Relatively, oblique vaginal septum will severely impair the patients' life quality and reproductive ability, with earlier onset and more complications. Besides, oblique vaginal septum can be divided into three subtypes: Type I, with a complete closure on obstructed-side vagina; Type II, patients with a small hole on the septum; Type II, with the presence of a fistula between two-sided cervixes. Type I patients tend to suffer from heavy dysmenorrhea and therefore are early diagnosed

Wang *et al. BMC Medical Imaging*   (2024) 24:347

Page 8 of 12
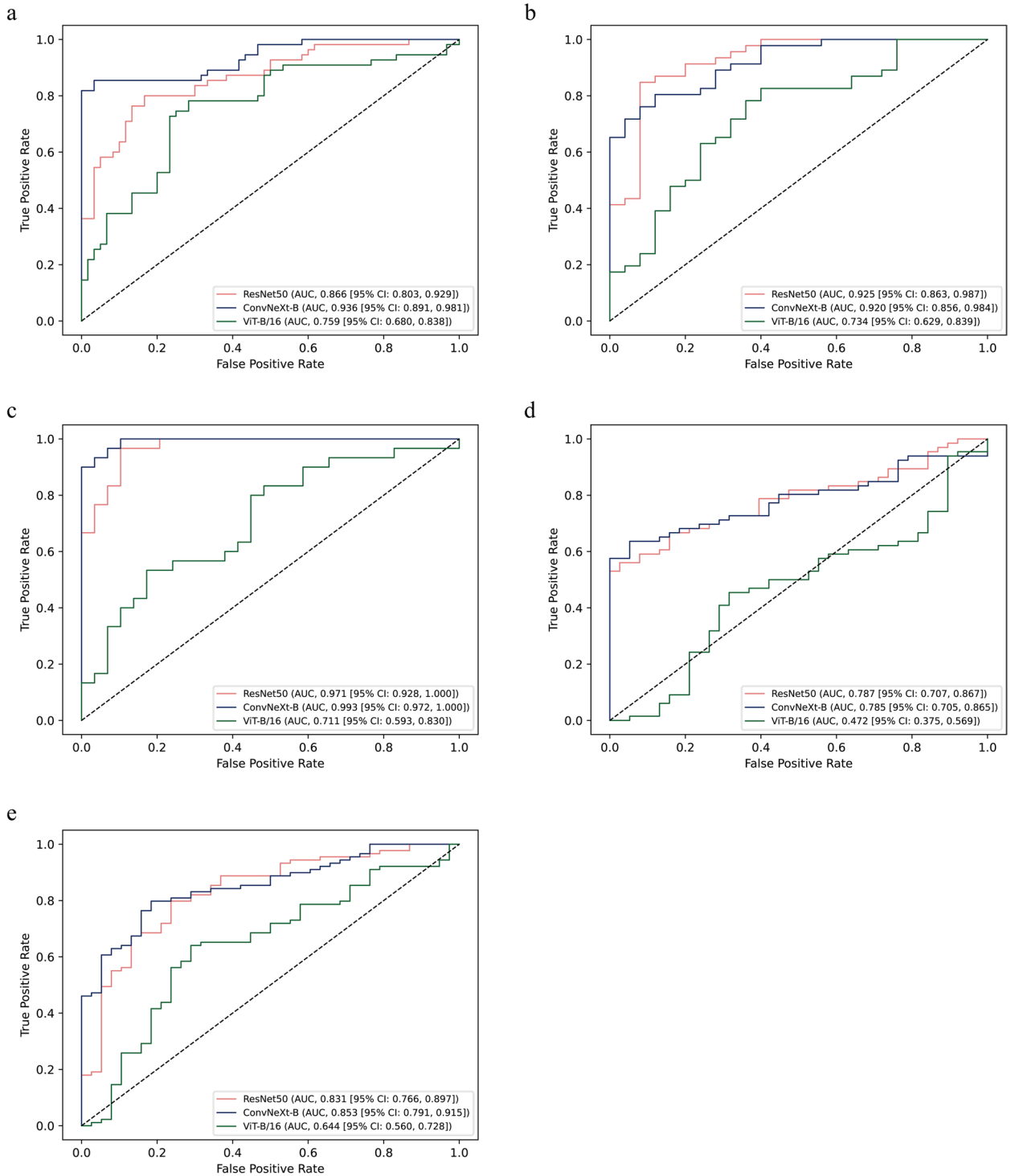
Fig. 4 ROC of DL models on validation set according to 5-fold cross-validation. **a**. ROC of cv1 on validation set; **b**. ROC of cv2 on validation set; **c**. ROC of cv3 on validation set; **d**. ROC of cv4 on validation set; **e**. ROC of cv5 on validation set. ROC, the receiver operating curve. DL, deep learning

Wang *et al. BMC Medical Imaging*   (2024) 24:347

Page 9 of 12

Table 3 Discriminative performance of case-based DL models according to 5-fold cross-validation

Model	Dataset	Accuracy [Variance, (95%CI)]	Sensitivity [Variance, (95%CI)]	Specificity [Variance, (95%CI)]	AUC [Variance, (95%CI)]
ResNet50	Training	1.000 (0.000, [1.000–1.000])	1.000 (0.000, [1.000–1.000])	1.000 (0.000, [1.000–1.000])	1.000 (0.000, [1.000–1.000])
ResNet50	Validation	0.772 (0.005, [0.534–0.988])	0.808 (0.030, [0.631–0.960])	0.709 (0.041, [0.505–0.905])	0.842 (0.004, [0.639–0.997])
ConvNeXt-B	Training	1.000 (0.000, [1.000–1.000])	1.000 (0.000, [1.000–1.000])	1.000 (0.000, [1.000–1.000])	1.000 (0.000, [1.000–1.000])
ConvNeXt-B	Validation	**0.817 (0.004, [0.598–0.994])**	**0.817 (0.016, [0.632–0.984])**	**0.811 (0.017, [0.622–0.979])**	**0.897 (0.004, [0.734–1.000])**
ViT_B/16	Training	1.000 (0.000, [1.000–1.000])	1.000 (0.000, [1.000–1.000])	1.000 (0.000, [1.000–1.000])	1.000 (0.000, [1.000–1.000])
ViT_B/16	Validation	0.668 (0.014, [0.407–0.920])	0.725 (0.040, [0.522–0.925])	0.572 (0.024, [0.304–0.831])	0.681 (0.030, [0.434–0.908])

Abbreviation: *DL* deep learning, *95%CI* 95% confidence interval, *AUC* area under the curve

Table 4 Comparison of the discriminative performance of case-based DL models in validation set

Model 1	Model 2	P_value (Accuracy)	P_value (Sensitivity)	P_value (Specificity)	P_value (AUC)
ResNet50	ViT_B/16	**0.016**	0.151	**0.032**	**< 0.01**
ConvNeXt-B	ViT_B/16	**0.036**	0.249	**< 0.01**	**< 0.01**
ResNet50	ConvNeXt-B	0.834	0.753	0.548	0.841

Abbreviation: *DL* deep learning, *AUC* area under the curve

while type II and type III patients may encounter a delayed diagnosis for incomplete obstruction [5, 8].

Auxiliary imaging like ultrasonography and magnetic resonance imaging has opened a novel era of non-invasive diagnosis of congenital female reproductive tract anomalies [28]. Nowadays, for the integrated advantages, 3D-US is a clinically frequently selected diagnostic imaging approach for female reproductive malformations. However, due to factors like confusing manifestations and complex anatomical structures, sonographers are confronted with a series of difficulties when using 3D-US to classify longitudinal and oblique vaginal septa. For instance, patients with longitudinal vaginal septum can be correlated with uterus malformations, similar to patients with oblique vaginal septum [29–31]. In addition, patients with type II/III oblique vaginal septum can discharge menstrual blood through the small septal hole or the cervix fistula, thus complaining alleviated obstructive symptoms, which are likely to be confused with longitudinal vaginal septum [5]. Simultaneously, both the
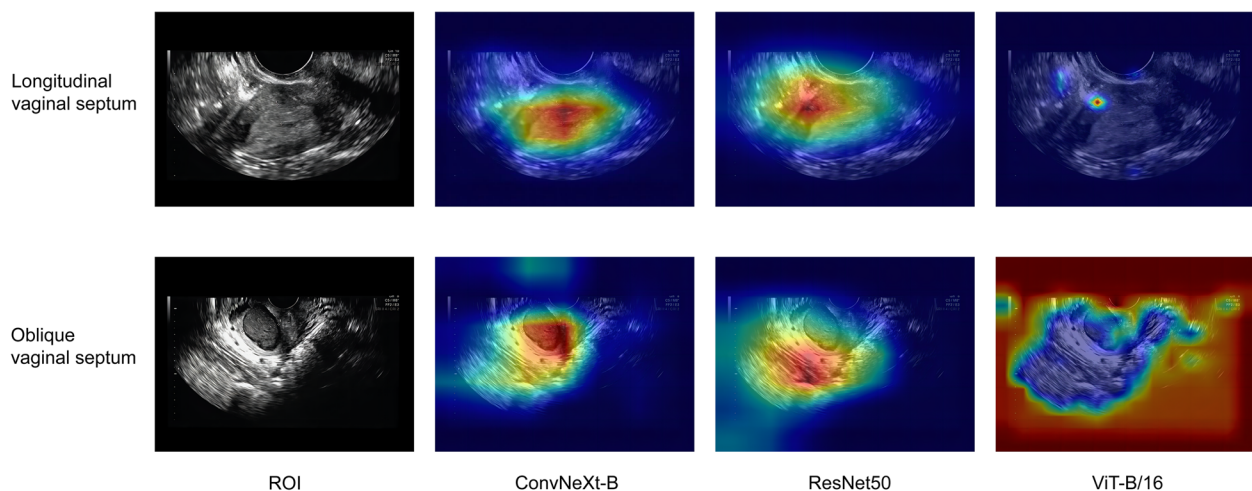
Fig. 5 Feature map of representative images on validation set according to 5-fold cross-validation. An image of longitudinal vaginal septum and an image of oblique vaginal septum were randomly selected to show the focused region of ConvNeXt-B, ResNet50 and ViT-B/16. DL, deep learning

Wang *et al. BMC Medical Imaging*   (2024) 24:347

Page 10 of 12

longitudinal and the oblique vaginal septa derive from alternative embryologic development of müllerian duct [29]. Clinicians and sonographers who lack the knowledge of embryology and mullerian duct anomalies and lack clinical exposure to related cases, may fail to distinguish longitudinal and oblique vaginal septa. As is reported in the literature and presented in Table.S2, oblique vaginal septum started to disturb the females after menarche [5], when transabdominal ultrasound imaging may be more suitable for examination but may produce unclear images if the patients were with visceral obesity or did not cooperate well. Moreover, without pre-existing indicative clinical symptoms, even experienced sonographer may omit the vaginal anomalies when the ultrasonic probe can be pushed forward successfully. In addition, lower-level or local hospital are scarcely exposed to cases of vaginal anomalies and usually poorly equipped, where a reliable diagnostic tool may greatly help. Taking the above issues into consideration, it's imperative to promote the diagnostic performance of ultrasound imaging for discriminating longitudinal and oblique vaginal septa.

As is reported, compared to CNN-based models, the transformer-based networks perform well in various domains through the self-attention mechanism and global information processing capability [27]. However, in this study, we found that ViT-B/16 does not perform as well as CNN-based models in all metrics and takes the longest time for training to reach convergence. In contrast, ConvNeXt-B not only performed better than ViT-B/16 in term of accuracy, sensitivity, specificity and AUC, but also can converge quickly in the training process.

Transformer was initially used in natural language processing and later its ideas were introduced to the field of image analysis, which is more suitable for tasks of global dependency like text translation. In this study, we aimed to identify the vaginal anomalies of ultrasonographic images in a DL supportive manner. However, given that the ROI of the ultrasonographic images of longitudinal and oblique vaginal septum can't be in detail depicted manually, we almost took the whole picture as input data. Therefore, the extra noisy information may lead to the poor performance of the DL models, especially the transformer-based model (ViT-B/16). Notably, CNN-based DL models are designed to capture the local dependency, thus achieving satisfactory performance in the image recognition task. ConvNeXt-B can capture features of variable scales through optimizing the traditional pooling layers and prompt the performance of network by co-opting from ResNets and by combing with features of variable scales. As we can see, compared to other DL models, ConvNeXt-B can reach convergence and show relatively better capability of classification with only several epoch.

We further tried to combine the three DL models. Compared with ConvNeXt-B, which has the best performance as a single model, the AUC of the combined model in both image-based and case-based recognition decreased, indicating that ensemble learning may be not necessary to improve the performance of ConvNeXt-B model.

To the best of our knowledge, this is the first time that deep learning models were used to identify longitudinal vaginal septum and oblique vaginal septum. ConvNeXt-B demonstrated the most outstanding performance with the AUC of 0.897, indicating its potential of accurate discriminative diagnosis of longitudinal vaginal septum and oblique vaginal septum in clinical application.

But there are still some limitations. The sample size seems relatively insufficient due to the low incidence of longitudinal and oblique vaginal septum. Besides, ultrasound images with normal vagina weren't enrolled to assess the ability of discerning vaginal anomalies. Because in clinical practice, sonographers do not capture and store the images of healthy vagina unless they have a preliminary diagnosis of vaginal anomalies before ultrasound imaging. What's more, we didn't enroll multicenter cases and images to establish and validate the DL models. In the future, multicenter data with larger sample-size will be collected to improve the robustness of the DL models. In addition, the inclusion criteria of image quality should be clearly defined and better conducted. Concerning the image segmentation, we utilized ITK-SNAP 4.0 to depict the ROI manually. If possible, computer-assisted image segmentation will be conducted to automatically extract the ROI in our following studies. Using the DL algorithms, we are aiming to achieve accurate classification of other MDAs, like uterus septum.

In conclusion, ConvNeXt-B shows good performance on the discriminating of longitudinal vaginal septum and oblique vaginal septum and may have promising application in DL aided ultrasonographic diagnostic network, further benefiting the establishment of appropriate clinical decision-making.

Abbreviations

3D-US	Three-dimensional ultrasonography
AUC	Area under the curve
CNN	Convolutional neural network
DL	Deep learning
MDA	Müllerian duct anomaly
MLP	Multilayer perceptron
ROC	Receiver operating characteristic
ROI	Region of interest
ViT	Vision transformer

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12880-024-01507-x.

Supplementary Material 1.

Wang *et al. BMC Medical Imaging*   (2024) 24:347

Page 11 of 12

Data availability
Relevant data supporting the findings of this study can be freely available from the corresponding author on reasonable requests, without commercial purposes and participants' confidentiality violation.

Declarations

Ethics approval and consent to participate
This study was approved by the Ethics Review Committee of Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology (approval number TJ-IRB20211225). Informed consent was waived for anonymized information of all individual participants included in this study.

Consent for publication
All listed authors have approved the manuscript for publication.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, No.1095 Jiefang Avenue, Wuhan, Hubei Province 430030, China. [2]Computer Center, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, No. 1095 Jiefang Avenue, Wuhan, Hubei Province 430030, China.

References
1. Ludwin A, Pfeifer SM. Reproductive surgery for müllerian anomalies: a review of progress in the last decade. Fertil Steril. 2019;112(3):408–16.
2. Sugi MD, Penna R, Jha P, Pōder L, Behr SC, Courtier J, Mok-Lin E, Rabban JT, Choi HH. Müllerian Duct anomalies: role in fertility and pregnancy. Radiographics. 2021;41(6):1857–75.
3. Venkata VD, Jamaluddin MFB, Goad J, Drury HR, Tadros MA, Lim R, Karakoti A, O'Sullivan R, Ius Y, Jaaback K, et al. Development and characterization of human fetal female reproductive tract organoids to understand Müllerian duct anomalies. Proc Natl Acad Sci U S A. 2022;119(30): e2118054119.
4. Pfeifer SM, Attaran M, Goldstein J, Lindheim SR, Petrozza JC, Rackow BW, Siegelman E, Troiano R, Winter T, Zuckerman A, et al. ASRM müllerian anomalies classification 2021. Fertil Steril. 2021;116(5):1238–52.
5. Jiang J, Yi S. Clinical features of 102 patients with different types of Herlyn-Werner-Wunderlich syndrome. Zhong Nan Da Xue Xue Bao Yi Xue Ban. 2023;48(4):550–6.
6. Borges AL, Sanha N, Pereira H, Martins A, Costa C. Herlyn-Werner-Wunderlich syndrome also known as obstructed hemivagina and ipsilateral renal anomaly: a case report and a comprehensive review of literature. Radiol Case Rep. 2023;18(8):2771–84.
7. Zhang Y, Liu Q. Diagnosis and treatment of 46 patients with oblique vaginal septum syndrome. J Obstet Gynaecol. 2022;42(8):3731–6.
8. Gai YH, Fan HL, Yan Y, Cai SF, Zhang YZ, Song RX, Song SL. Ultrasonic evaluation of congenital vaginal oblique septum syndrome: a study of 21 cases. Exp Ther Med. 2018;16(3):2066–70.
9. Ludwin A, Lindheim SR, Bhagavath B, Martins WP, Ludwin I. Longitudinal vaginal septum: a proposed classification and surgical management. Fertil Steril. 2020;114(4):899–901.
10. Bermejo C, Martínez-Ten P, Recio M, Ruiz-López L, Díaz D, Illescas T. Three-dimensional ultrasound and magnetic resonance imaging assessment of cervix and vagina in women with uterine malformations. Ultrasound Obstet Gynecol. 2014;43(3):336–45.
11. Berger A, Batzer F, Lev-Toaff A, Berry-Roberts C. Diagnostic imaging modalities for Müllerian anomalies: the case for a new gold standard. J Minim Invasive Gynecol. 2014;21(3):335–45.
12. Ergenoglu AM, Sahin Ç, Şimşek D, Akdemir A, Yeniel A, Yerli H, Sendag F. Comparison of three-dimensional ultrasound and magnetic resonance imaging diagnosis in surgically proven Müllerian duct anomaly cases. Eur J Obstet Gynecol Reprod Biol. 2016;197:22–6.
13. Grimbizis GF, Di Spiezio Sardo A, Saravelos SH, Gordts S, Exacoustos C, Van Schoubroeck D, Bermejo C, Amso NN, Nargund G, Timmerman D, et al. The Thessaloniki ESHRE/ESGE consensus on diagnosis of female genital anomalies. Hum Reprod. 2016;31(1):2–7.
14. Romanski PA, Bortoletto P, Pfeifer SM, Lindheim SR. An overview and video tutorial to the new interactive website for the American Society for Reproductive Medicine Müllerian anomalies classification 2021. Am J Obstet Gynecol. 2022;227(4):644–7.
15. Cheng C, Subedi J, Zhang A, Johnson G, Zhao X, Xu D, Guan X. Vaginoscopic incision of Oblique Vaginal Septum in adolescents with OHVIRA Syndrome. Sci Rep. 2019;9(1):20042.
16. Zhang J, Xie Y, Wu Q, Xia Y. Medical image classification using synergic deep learning. Med Image Anal. 2019;54:10–9.
17. Chen X, Wang X, Zhang K, Fung KM, Thai TC, Moore K, Mannel RS, Liu H, Zheng B, Qiu Y. Recent advances and clinical applications of deep learning in medical image analysis. Med Image Anal. 2022;79:102444.
18. Ramirez Zegarra R, Ghi T. Use of artificial intelligence and deep learning in fetal ultrasound imaging. Ultrasound Obstet Gynecol. 2023;62(2):185–94.
19. Xie HN, Wang N, He M, Zhang LH, Cai HM, Xian JB, Lin MF, Zheng J, Yang YZ. Using deep-learning algorithms to classify fetal brain ultrasound images as normal or abnormal. Ultrasound Obstet Gynecol. 2020;56(4):579–87.
20. Li Y, Cai P, Huang Y, Yu W, Liu Z, Liu P. Deep learning based detection and classification of fetal lip in ultrasound images. J Perinat Med. 2024;52(7):769–77.
21. Vafaeezadeh M, Behnam H, Gifani P. Ultrasound Image Analysis with Vision transformers-Review. Diagnostics (Basel). 2024;14(5):542.
22. Qiao S, Pang S, Sun Y, Luo G, Yin W, Zhao Y, Pan S, Lv Z. SPReCHD: four-Chamber Semantic Parsing Network for recognizing fetal congenital heart disease in Medical Metaverse. IEEE J Biomed Health Inf. 2024;28(6):3672–82.
23. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016;2015:770–8.
24. Liu Z, Mao H, Wu C, Feichtenhofer C, Darrell T, Xie S. A ConvNet for the 2020s. 2022 IEEE/CVF Conf Comput Vis Pattern Recognit (CVPR). 2022;11966:11976.
25. Wang Y, Lin Q, Sun ZJ, Jiang B, Hou B, Lu JJ, Zhu L, Feng F, Jin ZY, Lang JH. [Value of MRI in the pre-operative diagnosis and classification of oblique vaginal septum syndrome]. Zhonghua Fu Chan Ke Za Zhi. 2018;53(8):534–9.
26. Haverland R, Wasson M. Longitudinal vaginal septum: Accurate diagnosis and appropriate Surgical Management with Resection. J Minim Invasive Gynecol. 2020;27(6):1228–9.
27. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ArXiv e-prints*. 2020:arXiv :2010.11929.
28. Acién P, Acién M. Diagnostic imaging and cataloguing of female genital malformations. Insights Imaging. 2016;7(5):713–26.
29. De A, Jain A, Tripathi R, Nigam A. Complete uterine septum with cervical duplication and longitudinal vaginal septum: an anomaly supporting alternative Embryological Development. J Hum Reprod Sci. 2020;13(4):352–5.

30. Ludwin A. Ultrasound at the intersection of art, science and technology: complete septate uterus with longitudinal vaginal septum. Ultrasound Obstet Gynecol. 2021;57(3):512–4.

31. Mazzon I, Gerli S, Di Angelo Antonio S, Assorgi C, Villani V, Favilli A. The technique of vaginal septum as uterine septum: a New Approach for the Hysteroscopic treatment of vaginal septum. J Minim Invasive Gynecol. 2020;27(1):60–4.

Publisher's note

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