ORIGINAL RESEARCH ARTICLE

Assessing fetal lung maturity: Integration of ultrasound radiomics and deep learning

DOI: 10.29063/ajrh2025/v29i5s.7

Wanming Chen, Baohui Zeng, Xiaoyan Ling, Chen Chen, Jichuang Lai, Jianru Lin, Xihong Liu, Huien Zhou and Xinmin Guo**

Department of Ultrasound, Guangzhou Red Cross Hospital (Guangzhou Red Cross Hospital of Jinan University), Guangzhou, Guangdong 510220, China

***For Correspondence:** Email: *change90@126.com*; *397641515@qq.com*

Abstract

This study built a model to forecast the maturity of lungs by blending radiomics and deep learning methods. We examined ultrasound images from 263 pregnancies in the pregnancy stages. Utilizing the GE VOLUSON E8 system we captured images to extract and analyze radiomic features. These features were integrated with clinical data by means of deep learning algorithms such as DenseNet121 to enhance the accuracy of assessing fetal lung maturity. This combined model was validated by receiver operating characteristic (ROC) curve, calibration diagram, as well as decision curve analysis (DCA). We discovered that the accuracy and reliability of the diagnosis indicated that this method significantly improves the level of prediction of fetal lung maturity. This novel non-invasive diagnostic technology highlights the potential advantages of integrating diverse data sources to enhance prenatal care and infant health. The study lays groundwork, for validation and refinement of the model across various healthcare settings. *(Afr J Reprod Health 2025; 29 [5s]: 51-64).*

Keywords: Fetal; lung maturity; infant health; ultrasound radiomics

Résumé

Cette étude a permis d'élaborer un modèle de prédiction de la maturité pulmonaire en combinant la radiomique et des méthodes d'apprentissage profond. Nous avons examiné des images échographiques de 263 grossesses en cours de grossesse. Grâce au système GE VOLUSON E8, nous avons acquis des images pour extraire et analyser les caractéristiques radiomiques. Ces caractéristiques ont été intégrées aux données cliniques grâce à des algorithmes d'apprentissage profond tels que DenseNet121 afin d'améliorer la précision de l'évaluation de la maturité pulmonaire fœtale. Ce modèle combiné a été validé par la courbe ROC (Receiver Operating Characteristic), le diagramme d'étalonnage et l'analyse de la courbe de décision (DCA). Nous avons constaté que la précision et la fiabilité du diagnostic indiquaient que cette méthode améliorait significativement le niveau de prédiction de la maturité pulmonaire fœtale. Cette nouvelle technologie diagnostique non invasive met en évidence les avantages potentiels de l'intégration de diverses sources de données pour améliorer les soins prénatals et la santé infantile. Cette étude pose les bases de la validation et de l'affinement du modèle dans divers contextes de soins. *(Afr J Reprod Health 2025; 29 [5s]: 51-64).*

Mots-clés: Fœtal; maturité pulmonaire; santé infantile; radiomique échographique

Introduction

Abnormal or underdeveloped lung growth plays a role in respiratory issues among newborns.[1] The growth of the baby's lungs directly impacts their chances of survival. The crucial period for fetal lung development is from the 24 weeks to the 36 weeks of pregnancy.[2] In this period, the lungs are vulnerable to negative factors that can result in different abnormalities and impact the progress of fetal lung development.[3] The current primary method of assessing lung maturity in prenatal testing is through amniocentesis, a procedure that carries risks containing premature birth, early membrane rupture, and maternal bleeding.[4,5] Respiratory system abnormalities contain different conditions that influence the bronchi, lungs, or pulmonary blood vessels.

Prenatal ultrasound is a commonly utilized imaging method that uses sound waves to assess the fetus and its parts without being invasive. It is a tool to examine any problems in fetal development, with a focus on understanding the overall shape of the embryo and the important structure of the fetus.

Ultrasound imaging is portable and cost-effective and does not include any invasive procedures, making it a popular choice for diagnosing, screening, and managing various medical conditions.[6] Abundant scientific evidence indicate t that ultrasound can help monitor lung growth.[7,8] As prenatal ultrasound technology and skills have improved, it has become the method of choice for prenatal testing.[9,10] However, factors such as fluid levels, fetal bone formation, maternal weight issues and operator proficiency can sometimes result in errors or missed detections during ultrasound examinations.[11]

With advances in computer technology and the availability of massive ultrasound datasets, powerful artificial intelligence machine learning has been employed to automatically produce diagnoses by image analysis algorithms.[12,13] Deep learning is a branch of machine learning that adopts multi-layered neural networks to abstract and learn from representations of input data layer by layer, enabling the modeling of complex data structures and non-linear relationships. It has developed pattern recognition capabilities to help with detection, prediction, as well as classification tasks in medical image diagnosis, helping to surmount the limitations of the physician's visual assessment.[14,15] At present, deep learning has significant advantages in fields such as image recognition,[16] speech recognition,[17] and natural language processing.[18]

In ultrasound technology, the use of deep learning technology can assist clinicians to quickly and accurately assess clinical conditions, saving time in clinical diagnosis and treatment, and promoting the overall efficiency of ultrasound examination. This technology is currently being used for ultrasound diagnosis of obstetric, breast, heart, ovarian and liver diseases, including detection of fetal lung disease. It offers new tools and opportunities for diagnosing and treating various pediatric diseases.[19-21] Recently, with the advancement of medical diagnostic technology, the early detection of fetal health has been paid increasing attention. Literature review shows that in Spain (Quantitative Fetal Lung Maturity, QFLM), commercial software is used to conduct quantitative texture analysis of fetal lung to study the texture characteristics of ultrasound images, which preliminarily proves that the texture analysis of ultrasound images can indicate fetal lung maturity linked to gestational age.[22,23] Nevertheless, the implementation of artificial intelligence deep learning algorithms to assess fetal lung ultrasound images to predict fetal lung abnormalities is still restricted, and its value has not been extensively proven.[24,25]

Current techniques for lung ultrasound image analysis face two main challenges: the limited accuracy of traditional machine learning methods and the short of clarity of deep learning methods.[26-29] Traditional machine learning relies on feature extraction, which is not only time-consuming, but also subject to operator expertise, resulting in diminished accuracy when handling complex or uncommon situations.[30,31] Instead, deep learning proves the accuracy of image analysis. However, its opaque decision-making process poses challenges to interpretation, thus limiting its credibility and trustworthiness in clinical settings.[32,33] Therefore, strategies that provide clarity and interpretability while maintaining high precision are needed to improve the diagnostic validity and reliability of fetal lung ultrasound images.

The objective of this study is to obtain ultrasound images of lung development in the second trimester, identify key features, create a database, and apply advanced artificial intelligence techniques to build feature-based models. Thereafter, we conducted an evaluation with the goal of creating a reliable predictive tool for identifying abnormal fetal lung development that could be applied in clinical practice.

Methods

From January 2021 to October 2023, the Guangzhou Red Cross Hospital performed a second trimester fetal lung ultrasound scan on 305 pregnant women with medical history, and obtained 305 fetal lung ultrasound images. After excluding 42 cases involving known congenital issues or chromosomal disorders, twin pregnancies, and other specified exclusions, a total of 263 singleton pregnancies (consisting of 263 fetal lung ultrasound images) were considered for this investigation. The mean age of the participants was 31 years old, and the gestational age ranged from 17 to nearly 28 weeks. Inclusion criteria were: (1) singleton pregnancy; (2) complete medical history; (3) gestational age ranged 17-27+6 weeks. Exclusion criteria included: (1) fetal congenital abnormalities or chromosomal disorders; (2) poor image quality.

Each woman's stage of pregnancy was calculated based on the date of her last menstrual period, and then adjusted using an early pregnancy ultrasound to assess crown hip length. Among the 263 mothers, 65 (25.0%) experienced problems such as gestational diabetes, gestational hypertension or preeclampsia during pregnancy, while the other 198 (75.0%) had a successful pregnancy.

Ultrasound images acquisition

A GE VOLUSON E8 (GE, USA) color ultrasound diagnostic system was adopted for ultrasound examination of pregnant women using a matrix array transducer (C1-5-D probe) with a frequency range of 27 MHz and a center frequency of 3.5 MHz. Ultrasound scanning was conducted following the International Society of Ultrasound in Obstetrics and Gynecology (ISUOG) practice guidelines.[34] The manipulation and acquisition of fetal lung ultrasound images was done by two physicians, one with over 10 years of experience in obstetrics ultrasound (Physician 1) and another with 5 years of experience (Physician 2). Subsequently, fetal lung ultrasound images were separated into a training set and a validation set in random.

Fetal lung ultrasound images collected must meet the following criteria: (1) On the four-compartment view of the heart, the probe was adjusted to guarantee that at least one side of the fetal lung did not have the shadow effect of the fetal ribs; (2) The image was adjusted for every pregnant woman and fetus, containing depth, gain, frequency, time gain compensation as well as the use of harmonics (for obese pregnant women) to ensure the best image quality; (3) When saved, no color Doppler, arrows, or measurement values were retained on the image, and all images were collected and stored in DICOM format.[35]

Radiomics procedure

The program diagram of ROI segmentation, feature extraction, feature selection as well as clinical use of radiomics in fetal lung analysis are shown in Figure 1.

Feature extraction

In this study, we meticulously categorized and extracted handcrafted radiomic features, classifying them into three principal categories: geometric shape features, intensity features, and texture features.[36]

Geometric shape features are intended to capture the spatial geometry of tumors, which is vital for comprehending their growth patterns and potential invasiveness. Specifically, we calculated measurements of volume, surface area, surface, smoothness, roundness and various other characteristics related to shape. These parameters not outline the tumors shape and dimensions but also offer valuable information, about the intricacy and properties of tumor boundaries.

Intensity characteristics examine how the signal intensity is spread out among the image voxels by looking at statistical measures. We didn't just calculate statistics like average intensity middle value, variance, highest and lowest values; we also checked the consistency and diversity, in the tumor signals by analyzing skewness and kurtosis showing how the intensity is distributed within the tumor.

Texture characteristics use intricate second order and higher order techniques to examine patterns of voxel intensity and their spatial arrangement. We studied aspects of texture including image coarseness orientation, consistency and local variability. These characteristics offer a view of the intricate inner structures within the tumor, which are linked to gene expression and potential biological behavior. To capture these texture features, a variety of image analysis methods containing Gray Level Co-occurrence Matrix (GLCM) were applied for assessing the occurrence frequency of intensity pairs Gray Level Run Length Matrix (GLRLM) for measuring the length of consecutive runs with identical intensities Gray Level Size Zone Matrix (GLSZM) for quantifying continuous pixel area sizes with identical intensities and Neighbourhood Gray Tone Difference Matrix (NGTDM) for gauging intensity variation complexity, in local regions of the image.

To ensure the reliability along with reproducibility of the features we extracted, we utilized the pyradiomics tool (version 3.0.1) in adherence to the guidelines established by the Imaging Biomarker Standardization Initiative (IBSI). These guidelines are designed to promote consistency across studies and imaging devices providing a standardized and validated approach to the process of extracting radiomic features. This helps make our findings comparable and trustworthy. Additionally, we made adjustments to

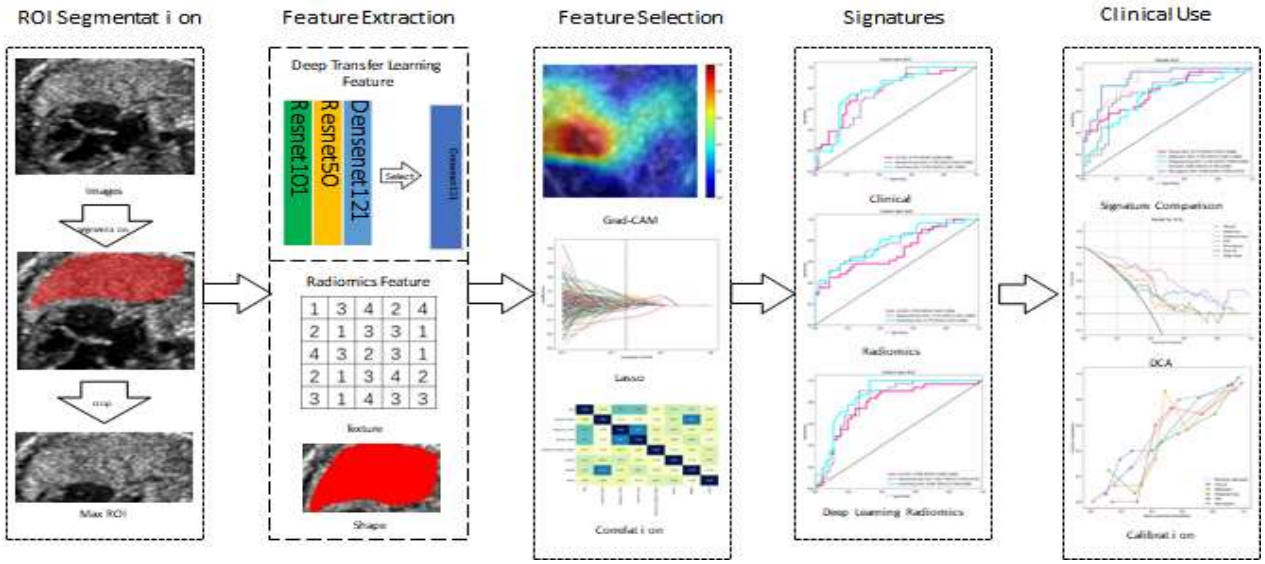

Figure 1: Workflow diagram for ROI segmentation, feature extraction, feature selection, and clinical application in fetal lung analysis

each subregion of the ultrasound images to explain differences, in patients and tumor characteristics ensuring that the extracted features accurately capture the distinctive attributes of each individual tumor.

Feature selection

During the feature selection phase of our research, we followed a stepped approach. Initially, we used Z-scores to standardize features. The p-value calculation based on T-test was then used, and features with a P-value less than 0.05 were selected for retention. Next, we examined features by means of Pearsons' correlation coefficient and decided to keep only one feature showing a correlation above 0.9. This involved employing a recursive deletion strategy to eliminate redundancy. Lastly to further refine our signatures feature set we applied LASSO regression to reduce the influence of irrelevant features. The optimal regularization parameter λ, for this process, was determined through 10-fold cross-validation.

Radiomics model

After Lasso-based feature refinement, we evaluated risk using two types of models: linear models (LR) and tree-based models (RandomForest, ExtraTrees). For model hyperparameter tuning, 5-fold cross-validation was adopted on the training set, while Gridsearch algorithm was adopted for hyperparameter optimization. The model parameters with the best median performance were included for the final model training.

Deep learning procedure

Crop ROI: In our approach we chose the slice with the Region of Interest (ROI) as the major image for each patient. To simplify things and reduce background interference, in our analysis process we kept only the smallest rectangular area covering the ROI. This rectangle was then enlarged by 10 pixels based on studies emphasizing the importance of areas surrounding tumors.

Data Augmentation: We adjusted the intensity distribution evenly across the RGB channels by normalizing the images using Z scores. These adjusted images were used as inputs for our model. In training, real-time data enhancement techniques containing cropping, horizontal flipping, as well as vertical flipping were employed. When processing test images, we only focused on normalization.

Deep learning signature

Transfer Learning: We investigated how known networks, like ResNet50 ResNet101 and DenseNet121 can improve the effectiveness of conventional CNN based models. Furthermore, we

compared these models to determine the algorithm for our particular research needs.

Hyper Parameters: In our research we used transfer learning to make sure that the model works well with groups of patients who have varying characteristics. This method included starting the model with learned weights, from the ImageNet database to help it perform better with various types of data. One important part of our method was carefully adjusting the learning rate to improve how well the model works across datasets. To do this we used a strategy called cosine decay learning rate, which is explained below:

$$\eta t = \eta i + (\eta — \eta i\ Tcur\ )\ (1 + \cos(\quad)) \quad (1)$$

min 1 max min Ti

The notation ηi = 0 means the minimum learning rate, while ηi = 0.01 means the maximum learning rate. The termTi = 30 represents the number of epochs in the iterative training process. Other essential hyperparameters include the use of Stochastic Gradient Descent (SGD) as a loss function of the optimizer and softmax cross entropy.

Deep Learning Signature: In our model, the output probabilities calculated by the CNN were defined as the Deep Learning Signature.

Deep Learning Radiomics Signature: In our research we found that the Densenet121 model implemented best among the learning models tested. We focused on extracting features from its 'features.norm5' layer, which has a total of 50,176 dimensions due to the models complexity. To address overfitting issues, with this high dimensionality we used Principal Component Analysis (PCA) to decrease these features to a more practical 512 dimensions.

To create the Deep Learning Radiomics (DLR)[37] signature we utilized a fusion algorithm that merges 512 dimensional deep learning characteristics with 1,562 dimensional radiomic features resulting in a total of 2,074 dimensional features. Following this we proceeded with a method, to the one employed in radiomics for selecting features and building models.

Clinical use

Combined Model: In order to make it more applicable in real world scenarios we conducted analyses on clinical features individually and then systematically. We pinpointed the important ones. Next, we combined these clinical features with the outcomes, from our advanced radiomics model. This led to the development of a Linear Regression (LR) model resulting in the formation of the Combined Signature. We effectively represented this signature using a nomogram.

Metrics: To measure how well our models perform we created Receiver Operating Characteristic (ROC) curves. We also looked at how the model is calibrated by analyzing calibration curves and using Hosmer Lemeshow analysis. Additionally, we used Decision Curve Analysis (DCA) to judge the value of our predictive models.

Statistical analysis

The statistical analysis was implemented by means of Python 3.7.12 with version 0.13.2 of the stats models library for statistical modeling. In developing the machine learning model, we adopted version 1.0.2 of the scikit learn library. Deep learning training was implemented on an NVIDIA 4090 GPU within the MONAI 0.8.1 and PyTorch 1.8.1 frameworks. We used a method to split the dataset, putting 70% of the data into the training group and the remaining 30% into the internal validation group. The fairness of this split was verified statistically as seen in Table 1 which reveals no variations in baseline traits, between the training and validation sets (all p-values > 0.05) suggesting an impartial separation of the dataset.

Ethical clearance

This study was consistent with the ethical standards of the 1964 Declaration of Helsinki and its later amendments, and was approved by the Ethics Committee of Guangzhou Red Cross Hospital and Tianhe District Maternal and Child Hospital of Guangzhou.

Results

Baseline characteristics

Table 1 provided a breakdown of the baseline features of the group. Factors such as age, week at delivery number of pregnancies ultrasound gestational week infant birth weight, number of deliveries, gender and Apgar score were all taken into account. These variables were compared in all participants, and significance was assessed using P-values in the training and experimental groups.

Table 1: Baseline characteristic of our cohort

Feature Name	ALL	Train	Test	P value
Age	30.95±4.11	31.11±4.28	30.57±3.68	0.341
Gestation Week	38.35±1.54	38.27±1.54	38.52±1.53	0.234
Pregnancy Times	1.85±1.01	1.88±0.98	1.77±1.06	0.4
Ultrasonic Gestation Week	23.55±2.03	23.61±2.09	23.40±1.89	0.463
Weight	3084.20±424.49	3065.25±421.53	3127.99±430.84	0.279
Delivery Times		-		0.275
0	2 (0.78)	1 (0.56)	1 (1.30)	-
1	158 (61.96)	106 (59.55)	52 (67.53)	-
2	89 (34.90)	68 (38.20)	21 (27.27)	-
3	6 (2.35)	3 (1.69)	3 (3.90)	-
Gender		-		0.105
0	124 (48.63)	93 (52.25)	31 (40.26)	-
1	131 (51.37)	85 (47.75)	46 (59.74)	-
Apgar score		-		0.103
0	237 (92.94)	169 (94.94)	68 (88.31)	-
1	18 (7.06)	9 (5.06)	9 (11.69)	-

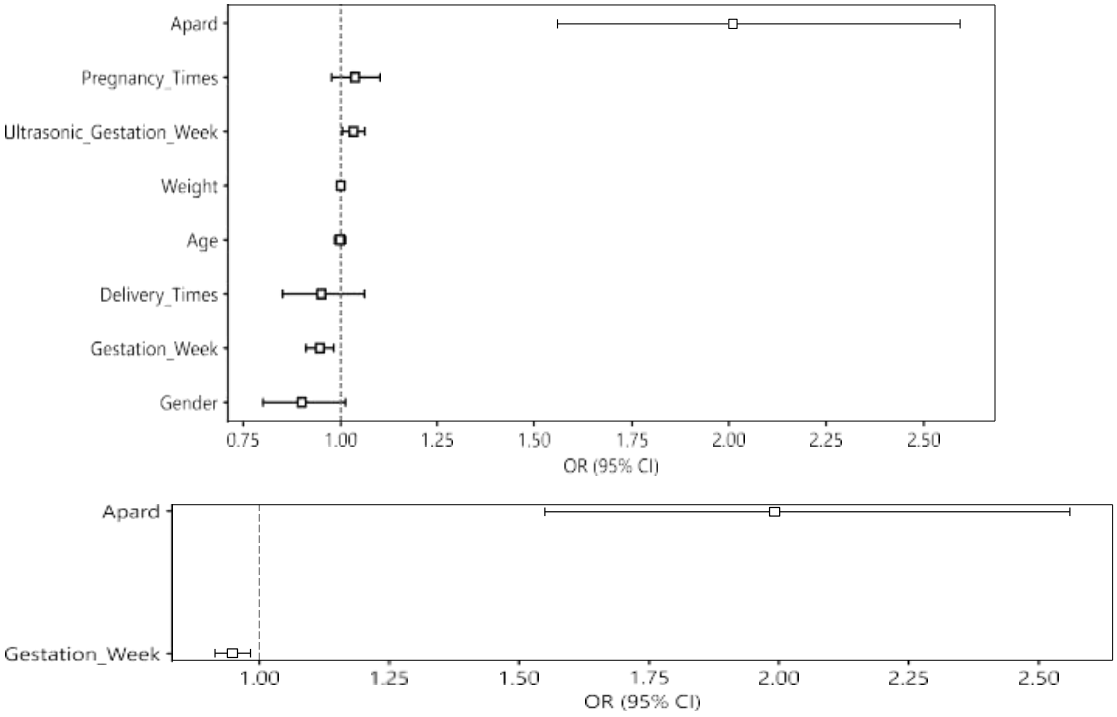

Figure 2: (a) OR of clinical features in univariable analysis. (b) OR of clinical features in multivariable analysis

Clinical features analysis

Univariable and multivariable Analysis: Each clinical characteristic was thoroughly examined through an individual analysis to determine the Odds Ratio (OR) and corresponding p values. Characteristics with a p value below 0.05 were deemed significant and included in the creation of the nomogram. Apart from the analysis a multivariable examination was carried out to adjust for potential influencing factors and identify the unique impacts of each clinical characteristic. The detailed process of this step analysis can be discovered in Supplementary Material 1A.

The following table provides a summary of both the individual and multivariable analyses presenting the

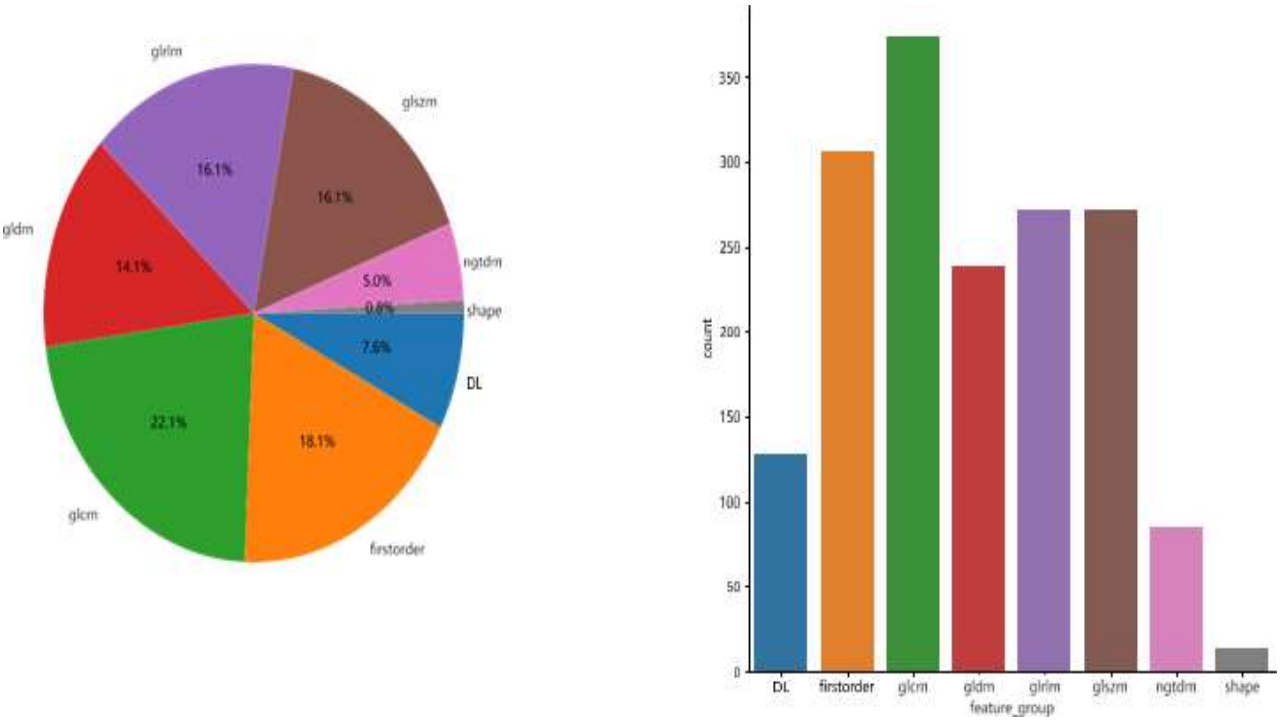

Figure 3: Number and ratio of handcrafted features

OR 95% Confidence Interval (CI) range and p value for each clinical characteristic. This table plays a role, in understanding the clinical significance of each variable.

Figure 2 (a) and Figure 2 (b) visually depict the Odds Ratios (ORs) for clinical features derived from univariate and multivariate analyses, respectively. These figures illustrate the comparative strength of association that each clinical feature has with the outcome, which in this study turned out to be the presence of a particular medical condition.

Deep learning radiomics signature

Handcrafted Features Statistics:[38] We carefully collected a wide range of 1,562 manually crafted radiomic characteristics, from the ultrasound pictures. These characteristics were categorized into shape, basic statistical properties and texture. Specifically, we outlined the process of acquiring 360 basic statistical properties, 14 shape attributes and a diverse assortment of texture features.

Figure 3 illustrates the breakdown and ratio of manually created features. This visual depiction highlights the makeup and equilibrium of the feature collection offering a glimpse into the intricacy of the dataset and the possible impact of each feature category.

Selecting radiomics features with lasso method

To choose the effective features we used LassoCV in conjunction, with a 10-fold cross validation approach. This method helped us to simplify the data and pinpoint the features that greatly impact model accuracy.

Figure 4 illustrated the results of the LassoCV feature selection process. It presented the Lasso path showing how the coefficients of features change with the regularization parameter, the squared error (MSE) from cross-validation, and the weights of the chosen radiomic features. These visuals provide an insight into how features are selected and highlight their relative importance.

Metric results for deep learning radiomics signature

In terms of evaluation metrics, we assessed the ExtraTrees model using measures such as accuracy, AUC, sensitivity, specificity positive predictive value (PPV) along with negative predictive value

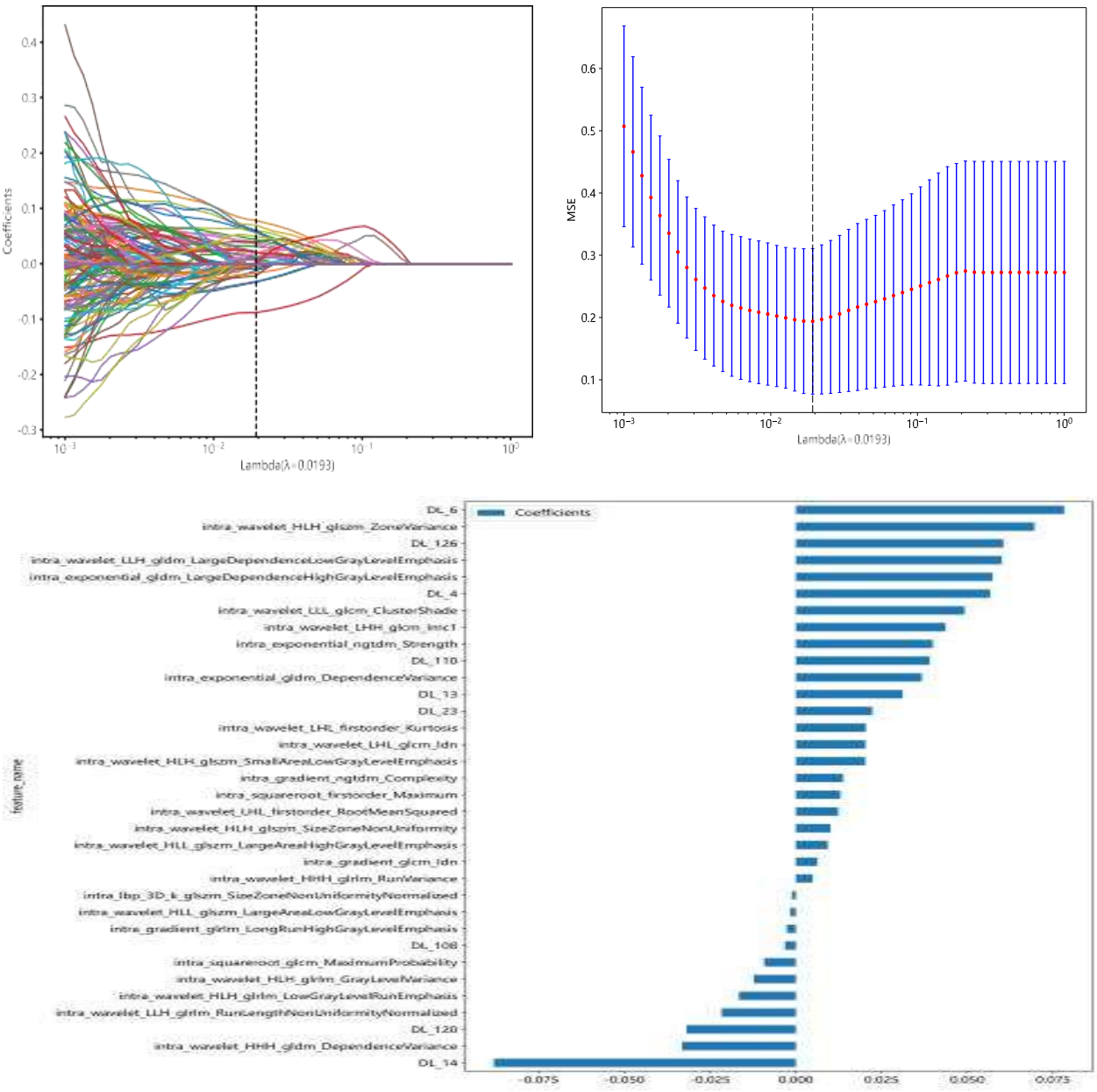

Figure 4: LassoCV feature selection process visuals

(NPV). The model showed predictive performance with high AUC scores, in both sets.

Table 2 provided a detailed overview of the performance metrics for the Deep Learning Radiomics Signature, comparing the results in different machine learning models within the training and testing cohorts. This comprehensive table emphasized the effectiveness of the ExtraTrees model, particularly in testing set, suggesting its suitability for practical diagnostic use.

ROC results for deep learning signature of different model

The ROC curves[39] in Figure 5 offer a comparative perspective on model performance among different data cohorts. They serve as an essential tool in evaluating and contrasting the true positive rate

Table 2: Metric results for Deep Learning Radiomics Signature

Model Name	Cohort	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Accuracy
LR	train	0.973	0.954 - 0.993	0.950	0.907	0.838	0.973	0.921
	test	0.788	0.682 - 0.893	0.806	0.674	0.625	0.838	0.727
Random Forest	train	0.947	0.917 - 0.976	0.900	0.847	0.750	0.943	0.865
	test	0.821	0.726 - 0.916	0.871	0.739	0.692	0.895	0.792
Extra Trees	train	0.914	0.874 - 0.954	0.750	0.890	0.776	0.875	0.843
	test	0.865	0.784 - 0.946	0.968	0.674	0.667	0.969	0.792

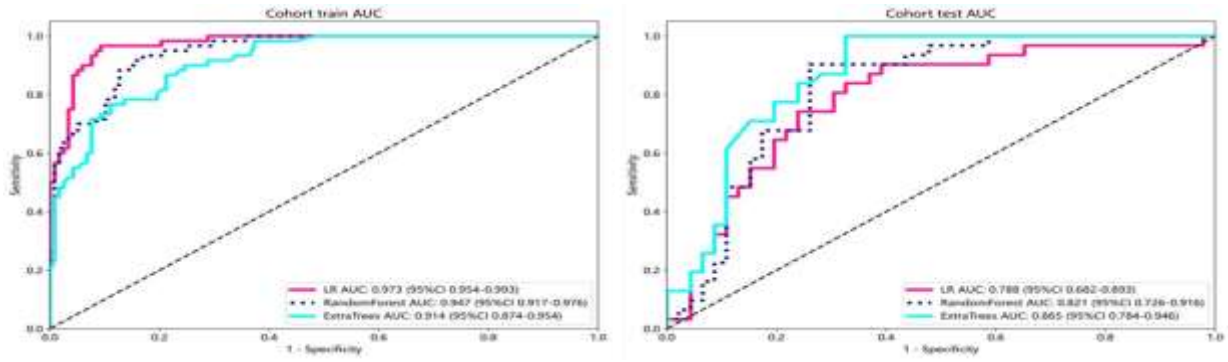

Figure 5: ROC results for Deep Learning Signature of different model

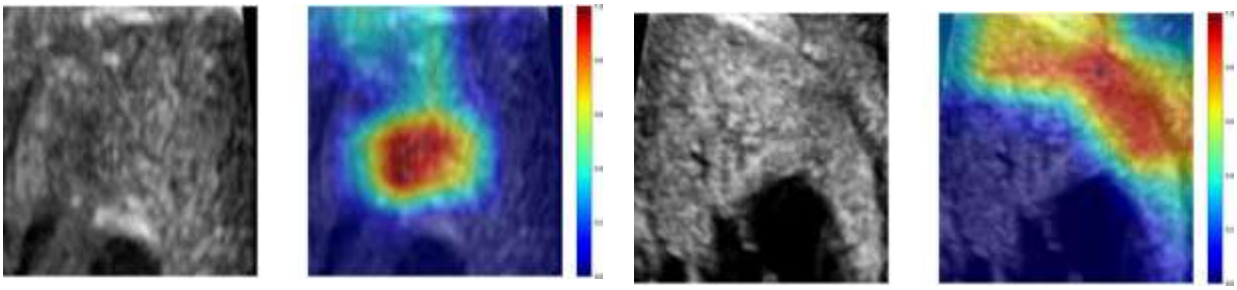

Figure 6: Grad-CAM visualizations for typical samples

Table 3: Metrics on different signature

Signature	Cohort	Accuracy	AUC 95% CI	Sensitivity	Specificity	PPV	NPV
Clinical	Train	0.781	0.8060.7352 - 0.8762	0.567	0.890	0.723	0.802
Radiomics	Train	0.787	0.8800.8296 - 0.9297	0.867	0.746	0.634	0.917
Deep Learning	Train	0.770	0.8430.7781 - 0.9080	0.800	0.754	0.623	0.881
DLR	Train	0.843	0.9140.8738 - 0.9537	0.750	0.890	0.776	0.875
Nomogram	Train	0.831	0.9280.8918 - 0.9635	0.867	0.814	0.703	0.923
Clinical	Test	0.727	0.7790.6727 - 0.8855	0.548	0.848	0.708	0.736
Radiomics	Test	0.753	0.7850.6809 - 0.8885	0.710	0.783	0.687	0.800
Deep Learning	Test	0.688	0.7280.6094 - 0.8457	0.710	0.674	0.595	0.775
DLR	Test	0.792	0.8650.7838 - 0.9463	0.968	0.674	0.667	0.969
Nomogram	Test	0.805	0.8910.8221 - 0.9605	0.774	0.826	0.750	0.844

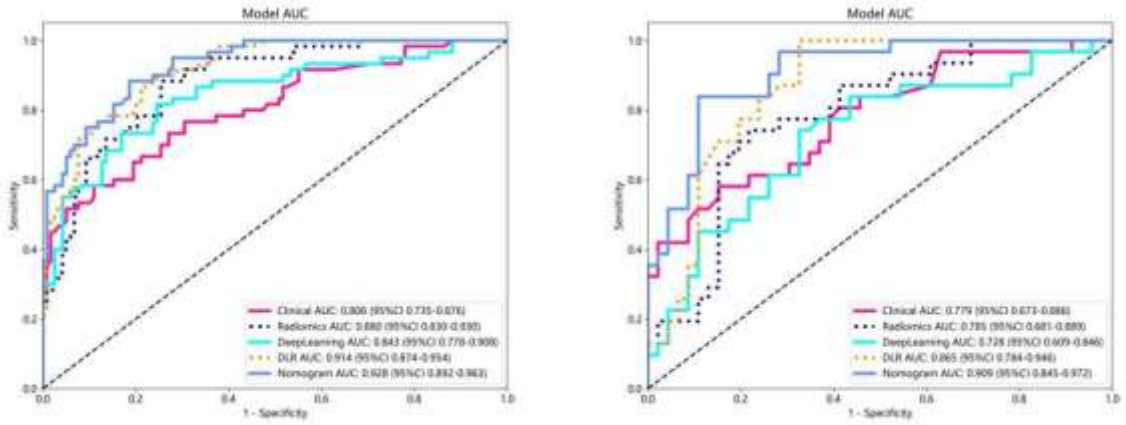

Figure 7: Receiver Operating Characteristic (ROC) curves for different signatures

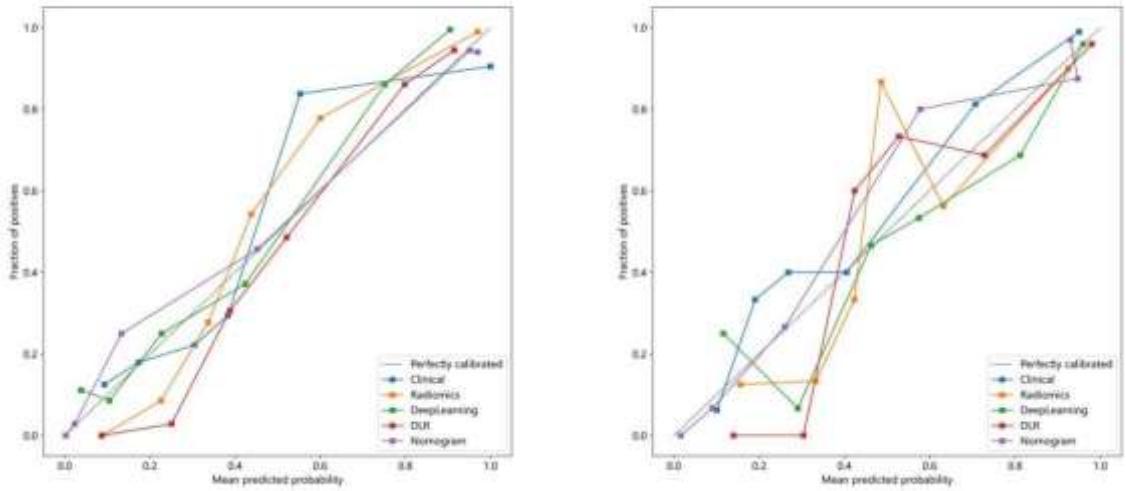

Figure 8: Calibration curves for different signatures

(sensitivity) against the false positive rate (1-specificity) across a range of thresholds.

Grad-CAM analysis

The utility of Grad-CAM[40] was to visually represent the focus areas of the deep learning model on the input images, effectively highlighting which regions were most influential in the predictive analysis. Figure 6 displays these visualizations for two representative samples, showcasing the model's attention and inference patterns.

Clinical application and evaluation

Our research went beyond evaluating individual models; we also delved into the blending of various modeling techniques. Through the amalgamation of Clinical, Radiomics and Deep Learning models to make a signature we witnessed a significant improvement, in performance measures.

In the evaluation of the integrated model, as depicted in Table 3 and Figure 7, the DLR model that incorporates deep learning and radiomics showed AUC values in both data sets compared to

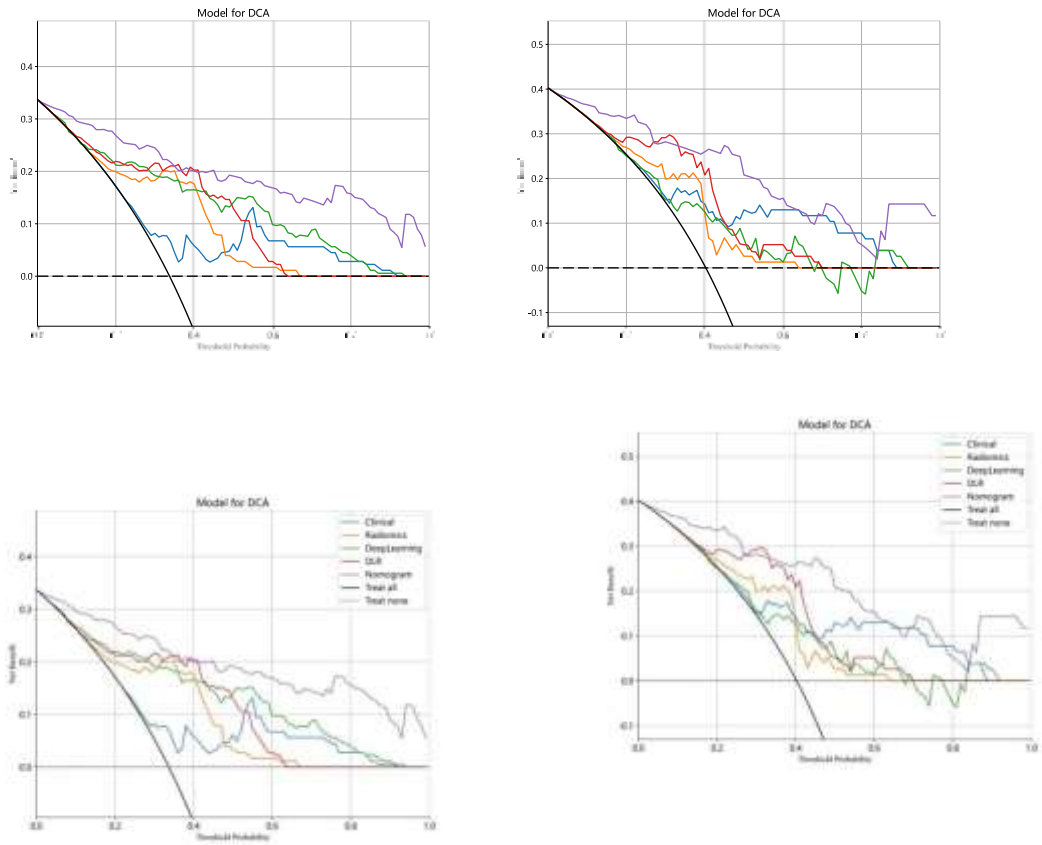

Figure 9: Different signatures' decision curve on test cohort

the standalone model. Nomograms incorporating features into DLR models obtained the best performance indicators, showing that combining different data inputs can greatly improve prediction accuracy.

Tuning models and ensuring accuracy

We improved our models further by conducting Hosmer Lemeshow (HL) tests to evaluate the calibration of these models. This test is important for checking if the predicted probabilities align closely with the outcomes. A calibrated model shows a higher HL statistic indicating accurate predictions that closely reflect real-world events. Figure 8 showcases the calibration curves providing a representation of this correlation, within our models.

Decision curve analysis (DCA)

To quantify the clinical usefulness of our predictive models, we performed DCA[41] for both the training and testing datasets, as depicted in Figure 9. The analysis showcases the net benefit across a range of decision thresholds, underlining the practical utility of our models.

Discussion

This research applied a dataset comprising 263 instances of singleton pregnancies laying a strong groundwork for training and validating the model. By integrating learning with traditional radiology techniques my goal was to improve the clarity and precision of our predictions. Compared to approaches our method enables earlier forecasts of fetal lung maturity. In addition, by using machine learning methods containing LassoCV selection features and PCA dimensionality reduction,[42] we ensure the credibility and importance of our radiological features. Our comprehensive model exhibited diagnostic accuracy through robust ROC curves, calibration plots, as well as DCA. This noninvasive technique effectively minimizes the risks associated with amniocentesis procedures such as premature labor and maternal bleeding.

The findings from our study carry implications for clinical application. The exceptional accuracy and reliability of our model imply its potential as a standard tool for routine prenatal screening potentially reducing the necessity for invasive interventions. The precise evaluation of lung maturity using solely ultrasound imaging is particularly beneficial in settings with limited access to advanced diagnostic resources. Additionally incorporating machine learning can augment clinicians' diagnostic abilities leading to timely and precise decision-making processes. By harnessing features and deep learning technologies we can extract valuable insights, from ultrasound images that enhance the predictive capacity of the model. Furthermore, utilizing a dataset from a single institution guarantees the uniformity of data quality and collection techniques.

However, this study has its limitations. The data was collected from one medical facility potentially restricting the broader applicability of the results. Future studies should confirm the model's effectiveness across centers and diverse populations to ensure its reliability and widespread usefulness. Despite efforts such as using Grad CAM to improve interpretability the opaque nature of learning models poses a challenge for their clinical implementation.

Our findings build on previous research regarding the application of radiomics as well as deep learning in medical imaging. While previous studies have emphasized the potential of radionics in evaluating lung maturity our research takes it a step further by enhancing predictive accuracy through deep learning integration.[4,5,7-10] Unlike machine learning approaches that rely on manual feature extraction, our approach leverages the automatic feature learning capabilities of deep learning to achieve more efficient and precise results.

We developed a predictive model for assessing fetal lung maturity, leveraging radiomics and deep learning techniques. Through meticulous analysis, including univariable and multivariable assessments, we integrated clinical, radiomic, and deep learning features, achieving high diagnostic accuracy and good model calibration. The combined approach showed significant predictive performance improvements, especially when clinical data were incorporated. In short, our thorough examination ranges from statistical validation to incorporating intricate models and evaluating their effectiveness in real-world medical settings. This diverse methodology not strengthens the reliability of our results but also sets the stage for meaningful clinical implementation ultimately driving progress, in personalized healthcare.

Strengths and limitations

The main strength of our study was that we developed a predictive model for assessing fetal lung maturity, leveraging radiomics and deep learning techniques, which may improve prenatal care and outcomes for neonatal health. The main limitation of our study was the applicability of the model in diverse environments remained unclear.

Conclusion

Our research introduces a comprehensible model for early identification of potential fetal lung development issues with the potential to improve prenatal care and outcomes for neonatal health. Our future efforts will focus on validating the applicability of the model in diverse environments and further refining its interpretability and effectiveness.

Competing interests

The authors report no actual or potential conflicts of interest.

Contribution of authors

Chen WM, Zeng BH and Ling XY: conceived and designed the study. Chen C, Lai JC and Lin JR: collected and analysed the data. Liu XH, Zhou HE and Guo XM: prepared the manuscript. All authors mentioned in the article approved the manuscript.

References

1. Laube M and Thome UH. Y It Matters-Sex Differences in Fetal Lung Development. Biomolecules. 2022; 12(3):
2. Liu D, Jiang Q, Xu Z, Li L and Lyu G. Evaluating fetal lung development at various gestational weeks using two-dimensional shear wave elastography. Quant Imaging Med Surg. 2024; 14(8): 5373-5384.
3. Wall J and Coates A. Prenatal imaging and postnatal presentation, diagnosis and management of congenital lung malformations. Curr Opin Pediatr. 2014; 26(3): 315-319.

4. Gordon MC, Narula K, O'Shaughnessy R and Barth WH, Jr. Complications of third-trimester amniocentesis using continuous ultrasound guidance. Obstet Gynecol. 2002; 99(2): 255-259.
5. Stark CM, Smith RS, Lagrandeur RM, Batton DG and Lorenz RP. Need for urgent delivery after third-trimester amniocentesis. Obstet Gynecol. 2000; 95(1): 48-50.
6. Zaffino P, Moccia S, De Momi E and Spadea MF. A Review on Advances in Intra-operative Imaging for Surgery and Therapy: Imagining the Operating Room of the Future. Ann Biomed Eng. 2020; 48(8): 2171-2191.
7. Ahmed B and Konje JC. Fetal lung maturity assessment: A historic perspective and Non - invasive assessment using an automatic quantitative ultrasound analysis (a potentially useful clinical tool). Eur J Obstet Gynecol Reprod Biol. 2021; 258343-347.
8. Palacio M, Bonet-Carne E, Cobo T, Perez-Moreno A, Sabrià J, Richter J, Kacerovsky M, Jacobsson B, García-Posada RA, Bugatto F, Santisteve R, Vives À, Parra-Cordero M, Hernandez-Andrade E, Bartha JL, Carretero-Lucena P, Tan KL, Cruz-Martínez R, Burke M, Vavilala S, Iruretagoyena I, Delgado JL, Schenone M, Vilanova J, Botet F, Yeo GSH, Hyett J, Deprest J, Romero R and Gratacos E. Prediction of neonatal respiratory morbidity by quantitative ultrasound lung texture analysis: a multicenter study. Am J Obstet Gynecol. 2017; 217(2): 196.e191-196.e114.
9. Whitworth M, Bricker L and Mullan C. Ultrasound for fetal assessment in early pregnancy. Cochrane Database Syst Rev. 2015; 2015(7): Cd007058.
10. Dias T, Sairam S and Kumarasiri S. Ultrasound diagnosis of fetal renal abnormalities. Best Pract Res Clin Obstet Gynaecol. 2014; 28(3): 403-415.
11. Wataganara T, Ebrashy A, Aliyu LD, Moreira de Sa RA, Pooh R, Kurjak A, Sen C, Adra A and Stanojevic M. Fetal magnetic resonance imaging and ultrasound. J Perinat Med. 2016; 44(5): 533-542.
12. Seyer Cagatan A, Taiwo Mustapha M, Bagkur C, Sanlidag T and Ozsahin DU. An Alternative Diagnostic Method for C. neoformans: Preliminary Results of Deep-Learning Based Detection Model. Diagnostics (Basel). 2022; 13(1):
13. Chen J, Huang Q, Chen Y, Qian L and Yu C-SJA. Enhancing Nucleus Segmentation with HARU-Net: A Hybrid Attention Based Residual U-Blocks Network. 2023; abs/2308.03382
14. Jiang Y, Yang M, Wang S, Li X and Sun Y. Emerging role of deep learning-based artificial intelligence in tumor pathology. Cancer Commun (Lond). 2020; 40(4): 154-166.
15. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Cui C, Corrado G, Thrun S and Dean J. A guide to deep learning in healthcare. Nat Med. 2019; 25(1): 24-29.
16. Shen D, Wu G and Suk HI. Deep Learning in Medical Image Analysis. Annual review of biomedical engineering. 2017; 19221-248.
17. LeCun Y, Bengio Y and Hinton G. Deep learning. Nature. 2015; 521(7553): 436-444.
18. Otter DW, Medina JR and Kalita JK. A Survey of the Usages of Deep Learning for Natural Language Processing. IEEE transactions on neural networks and learning systems. 2021; 32(2): 604-624.
19. Zhou LQ, Wang JY, Yu SY, Wu GG, Wei Q, Deng YB, Wu XL, Cui XW and Dietrich CF. Artificial intelligence in medical imaging of the liver. World J Gastroenterol. 2019; 25(6): 672-682.
20. Hosny A, Parmar C, Quackenbush J, Schwartz LH and Aerts H. Artificial intelligence in radiology. Nat Rev Cancer. 2018; 18(8): 500-510.
21. Chen Z, Liu Z, Du M and Wang Z. Artificial Intelligence in Obstetric Ultrasound: An Update and Future Applications. Front Med (Lausanne). 2021; 8733468.
22. Moreno-Espinosa AL, Hawkins-Villarreal A, Coronado-Gutierrez D, Burgos-Artizzu XP, Martínez-Portilla RJ, Peña-Ramirez T, Gallo DM, Hansson SR, Gratacòs E and Palacio M. Prediction of Neonatal Respiratory Morbidity Assessed by Quantitative Ultrasound Lung Texture Analysis in Twin Pregnancies. J Clin Med. 2022; 11(16):
23. Chen P, Chen Y, Deng Y, Wang Y, He P, Lv X and Yu J. A preliminary study to quantitatively evaluate the development of maturation degree for fetal lung based on transfer learning deep model from ultrasound images. Int J Comput Assist Radiol Surg. 2020; 15(8): 1407-1415.
24. Keerthi G and Abirami MSJMTA. Intelligent diagnosis of fetal organs abnormal growth in ultrasound images using an ensemble CNN-TLFEM model. 2024; 8381167-81178.
25. Fiorentino MC, Villani FP, Di Cosmo M, Frontoni E and Moccia S. A review on deep-learning algorithms for fetal ultrasound-image analysis. Medical image analysis. 2023; 83102629.
26. Avena-Zampieri CL, Hutter J, Rutherford M, Milan A, Hall M, Egloff A, Lloyd DFA, Nanda S, Greenough A and Story L. Assessment of the fetal lungs in utero. Am J Obstet Gynecol MFM. 2022; 4(5): 100693.
27. Xiao S, Zhang J, Zhu Y, Zhang Z, Cao H, Xie M and Zhang L. Application and Progress of Artificial Intelligence in Fetal Ultrasound. J Clin Med. 2023; 12(9):
28. Xia TH, Tan M, Li JH, Wang JJ, Wu QQ and Kong DX. Establish a normal fetal lung gestational age grading model and explore the potential value of deep learning algorithms in fetal lung maturity evaluation. Chin Med J (Engl). 2021; 134(15): 1828-1837.
29. Rigatti SJ. Random Forest. Journal of insurance medicine (New York, NY). 2017; 47(1): 31-39.
30. Karypidis E, Mouslech SG, Skoulariki K and Gazis AJA. Comparison Analysis of Traditional Machine Learning and Deep Learning Techniques for Data and Image Classification. 2022; abs/2204.05983
31. Jiang Y, Luo J, Huang D, Liu Y and Li DD. Machine Learning Advances in Microbiology: A Review of Methods and Applications. Front Microbiol. 2022; 13925454.
32. Zhou SK, Greenspan H, Davatzikos C, Duncan JS, van Ginneken B, Madabhushi A, Prince JL, Rueckert D and Summers RM. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. Proceedings of the IEEE Institute of

Electrical and Electronics Engineers. 2021; 109(5): 820-838.
33. Singh A, Sengupta S and Lakshminarayanan V. Explainable Deep Learning Models in Medical Image Analysis. Journal of imaging. 2020; 6(6):
34. Safety statement, 2000. International Society of Ultrasound in Obstetrics and Gynecology (ISUOG). Ultrasound Obstet Gynecol. 2000; 16(6): 594-596.
35. Pianykh OS. Digital Imaging and Communications in Medicine (DICOM). In: 2012.
36. Banerjee V, Wang S, Drescher M, Russell R and Siddiqui MM. Radiogenomics influence on the future of prostate cancer risk stratification. Ther Adv Urol. 2022; 14175628722211253l7.
37. Li S, Liu J, Wang Z, Cao Z, Yang Y, Wang B, Xu S, Lu L, Iqbal Saripan M, Zhang X, Dong X and Wen DJRS. Application of PET/CT-based deep learning radiomics in head and neck cancer prognosis: a systematic review. 2022
38. Nanni L, Ghidoni S and Brahnam SJPR. Handcrafted vs. non-handcrafted features for computer vision classification. 2017; 71158-172.
39. Zweig MH and Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem. 1993; 39(4): 561-577.
40. Khodadadi Shoushtari F, Dehkordi ANV and Sina S. Quantitative and Visual Analysis of Data Augmentation and Hyperparameter Optimization in Deep Learning-Based Segmentation of Low-Grade Glioma Tumors Using Grad-CAM. Ann Biomed Eng. 2024; 52(5): 1359-1377.
41. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, Roobol MJ and Steyerberg EW. Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. Eur Urol. 2018; 74(6): 796-804.
42. Tibshirani RJJotrsssb-m. Regression Shrinkage and Selection via the Lasso. 1996; 58267-288.