

Prenatal Prediction of Neonatal Respiratory Morbidity: A Radiomics Method Based on Imbalanced Few-Shot Fetal Lung Ultrasound Images

Jing Jiao

Fudan University

Yanran Du

Fudan University

Xiaokang Li

Fudan University

Yi Guo (✉ guoyi@fudan.edu.cn)

Fudan University <https://orcid.org/0000-0002-7142-2871>

Yunyun Ren

Fudan University

Yuanyuan Wang

Fudan University

Research article

Keywords: neonatal respiratory distress syndrome, transient tachypnea, prenatal ultrasonic diagnosis, fetal lung ultrasound image, class imbalance, ensemble learning

Posted Date: June 28th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-654307/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: To develop a non-invasive method for the prenatal prediction of neonatal respiratory morbidity (NRM) by a novel radiomics method based on imbalanced few-shot fetal lung ultrasound images.

Methods: A total of 210 fetal lung ultrasound images were enrolled in this study, including 159 normal newborns and 51 NRM newborns. Fetal lungs were delineated as the region of interest (ROI), where radiomics features were designed and extracted. Integrating radiomics features selected and two clinical features, including gestational age (GA) and gestational diabetes mellitus (GDM), the prediction model was developed and evaluated. The modelling methods used were data augmentation, cost-sensitive learning, and ensemble learning. Furthermore, two methods, which embed data balancing into ensemble learning, were employed to address the problems of imbalance and few-shot simultaneously.

Results: Our model achieved sensitivity values of 0.82, specificity values of 0.84, accuracy values of 0.84 and area under the curve values of 0.87 in the test set. The radiomics features extracted from the ROIs at different locations within the lung region achieved similar classification performance outcomes.

Conclusion: The feature set we designed can efficiently and robustly describe fetal lungs for NRM prediction. RUSBoost shows excellent performance compared to state-of-the-art classifiers on the imbalanced few-shot dataset. The diagnostic efficacy of the model we developed is similar to that of several previous reports of amniocentesis and can serve as a non-invasive, precise evaluation tool for NRM prediction.

1 Background

Neonatal respiratory morbidity (NRM), mainly including respiratory distress syndrome (RDS) and transient tachypnea of the newborn (TTN), is a leading cause of morbidity and mortality in the preterm and early term [1]. The morbidity of NRM is correlated with fetal lung maturity [2]. Newborns with NRM are born with respiratory distress and even apnoea, which may lead to multiple complications, or even death. Glucocorticoids are used to treat fetuses at high risk of NRM to promote fetal lung maturation and can significantly reduce morbidity and mortality. However, recent studies have shown that glucocorticoid treatment has some side effects, such as short-term fetal heart rate variability (HRV) and fetal movements [3]. An accurate prenatal prediction of NRM is essential to avoid the overuse of glucocorticoids in normal fetuses.

Amniocentesis is an effective method for the prenatal prediction of NRM by assessing fetal lung maturity [4]. However, it is an invasive detection method with complicated and time-consuming operations and no uniform threshold for the prediction. Currently, amniocentesis is rarely used to make prenatal predictions. Instead, gestational age (GA) is usually assessed to make the prediction. Fetuses assessed to be born at 28-36.6 weeks are regarded as having a high risk of NRM because of fetal lung immaturity and will be treated with glucocorticoids. There is a high rate of false positives in view of NRM morbidity, which will

cause side effects in newborns. In this context, it is particularly important to develop an accurate and non-invasive method for the prenatal prediction of NRM.

Ultrasound is a non-radiation and non-invasive technology that is widely used in prenatal diagnosis. The use of fetal lung ultrasound images to predict NRM as alternative to amniocentesis has been considered a useful method in recent studies [5]. In a recent study, quantitative texture analysis of fetal lungs (quantusFLM) was used to predict NRM [6]. The study was based on the European population and no related study for Asian populations. Moreover, the feature set used in their study only includes textural features and GA. There is suggestive evidence that gestational diabetes mellitus (GDM) in pregnant women may have adverse effects on lung development [7][8]. On the other hand, due to low morbidity, NRM newborns, especially preterm and early-term newborns, are hard to obtain. The dataset for the study is usually imbalanced and few-shot. This phenomenon was not mentioned in their study. It is worth noting that imbalanced and few-shot datasets are common in clinical practice and will bring overfitting and bias, resulting in poor generalization for the classification model.

The purpose of this study was to develop a non-invasive method for the prenatal prediction of NRM based on the radiomics method with an imbalanced few-shot fetal lung ultrasound image dataset collected from Asian population. Fetal lungs were delineated as the region of interest (ROI), and radiomics features were designed and extracted from the ROI. Feature selection was performed to select representative radiomics features and combining with GA and GDM for modelling. The modelling method of data augmentation, cost-sensitive learning, ensemble learning, Random Under-Sampling with AdaBoost (RUSBoost) [9] and Synthetic Minority Oversampling Technique (SMOTE) with AdaBoost (SMOTEBoost) [10] were used to address the problems of imbalance and few-shot. Finally, the diagnostic efficacy of the model we developed was found to be similar to that of previous reports of amniocentesis.

2 Materials And Methods

2.1 Patients

From July 2018 to August 2019, a total of 261 fetal lung ultrasound images from 261 singleton pregnant women with GAs ranging from 28.0 to 38.6 weeks were collected from Obstetrics and Gynecology Hospital Affiliated to Fudan University, Shanghai, China. The flowchart for the study population is shown in **Figure 1**. Pregnant women who met the following criteria were enrolled in the study: 1) singleton pregnancy; 2) those with complete medical information who had undergone maternity examination and subsequent delivery in our hospital; 3) fetuses with no known congenital malformation or chromosomal abnormality; 4) those with no diabetes before pregnancy; and 5) those who had not been prescribed steroids before delivery. Finally, a total of 210 singleton pregnant women with 210 fetal lung ultrasound images were enrolled in our study and randomly divided into the training set and test set at a ratio of approximately 8:2. It is worth noting that we kept the same proportion of NRM and normal in both sets. The training set contains 167 images, of which 40 are NRM and 127 are normal. The test set contains 43 images, of which 11 are NRM and 32 are normal.

This study was approved by the Ethics Committee of Obstetrics and Gynecology Hospital Affiliated to Fudan University, Shanghai, China. All data were collected and used with the consent of the pregnant women.

2.2 Image acquisition and lung segmentation

All ultrasound images were obtained during routine prenatal ultrasound examinations within 72 hours before delivery and performed by a radiologist with over eight years of experience in obstetrics and gynaecology ultrasound imaging. Fetal lung image acquisition was achieved using a transverse view of the fetal thorax at the level of the 4-cardiac-chamber view. Considering that the fetus' right and left lungs were supposed to have the same development, we chose to adjust the probe to ensure that at least one side of the lungs was complete and had no obvious acoustic shadows from the fetal ribs. To obtain the best quality image, the settings were adjusted according to the condition of each pregnant woman and fetus, including depth, gain, frequency, time-gain compensation and harmonics in obese patients.

Since the lung region is homogeneous, the ROI selected is one part of the lung that contains only lung tissue and no vascular or rib shadows. The manual delineation of each fetal lung was delineated by one physician, which was reviewed and confirmed by another physician, both of whom were blinded to the medical histories of the pregnant women and neonatal outcomes. Two examples are shown in **Figure 2**.

2.3 Feature extraction and selection

The feature design is the basis for building a practical and generalizable classification model. For ultrasound fetal lung images, the feature set should reflect subtle texture information in the ROI of the image and independent of the ROI's size and location to provide a robust description for clinical use. With the requirement for the feature set, a series of radiomics features were designed based on the image greyscale and texture, including 16 greyscale histogram features, 60 texture features, and 304 wavelet features.

In addition, we used a priori clinical knowledge to improve the feature set's descriptive ability by adding two clinical features, GA and GDM, which are readily available and strongly correlated with NRM in relevant studies. The summary of the feature set is listed in **Table 1**, and the details of the features are as follows.

The feature selection method was used to select the most useful radiomics features as inputs of the classification model. We ranked feature importance to selected features by permuting out-of-bag data feature of random forest trees. If a feature is influential, permuting its values would influence the model error testing with out-of-bag data. The more important a feature is, the greater its influence will be [18].

2.4 Model building

The class imbalance and small dataset will lead to overfitting and classification bias. In this study, we designed and evaluate performance of common methods on our imbalanced and few-shot dataset. Here, the data augmentation method we used was Adaptive Synthetic (ADASYN) [19], which can generate minority class pseudo-samples concentrated at the classification boundary by linear interpolation. The cost-sensitive learning method used is the cost-sensitive support vector machine (SVM) [20], which improves the attention of the SVM to the minority classes. The ensemble learning method used was adaptive boosting (AdaBoost) [21], which can improve the generalization and classification performance of the model by combining weak base learners.

On the basis of AdaBoost, RUSBoost and SMOTEBoost, which embed data balancing into ensemble learning to address the problems of few-shot and imbalance simultaneously was also employed.

All classifier parameters were tuned with bootstrap 5-fold cross-validation, and the decision tree was employed as the base learner for AdaBoost, RUSBoost and SMOTEBoost.

2.5 Statistical analysis

Descriptive statistics are summarized as the mean standard deviation (mean std). Univariate analyses were performed on each feature of the training set using the t -test for 380 continuous radiomics features and the test for two categorical clinical features. A p value <0.05 indicated a significant difference.

The metrics used to evaluate the classification performance of the model are the accuracy (ACC), the area under the curve (AUC), the sensitivity (SENS) and the specificity (SPEC).

All methods were performed with MATLAB R2019b (MathWorks, Inc., Natick, MA, USA). The image processing toolbox and machine learning toolbox were applied in feature extraction and model building.

3 Result

3.1 Patient characteristics

A summary of the characteristics of the training set and test set is listed in **Table 2**. The imbalance ratio between the number of normal and NRM was close to 3:1. There was a significant difference (p value <0.005) in both GA and GDM between NRM and normal controls, which is the statistical basis for using GA and GDM as clinical features. Moreover, there was a significant difference (p value <0.0001) in birth weight between the two groups.

3.2 Univariate analysis and feature selection

Univariate analysis was performed on the training set. The results show that 32 of all 380 radiomics features were highly correlated with NRM (p value <0.05).

The feature selection method was used to select the most useful features for modelling. The final 10 features with the highest feature's importance score were selected. The feature names and descriptive statistics of the 10 radiomics features selected are listed in **Table 3**. **Figure 3** shows the box plots of the top 3 features with a high correlation between the normal and NRM fetal lung ultrasound images of the 10 selected features. Although there are significant differences in the means, the standard deviations overlap, making the classification task difficult and requires a more powerful multivariate classification method.

3.3 Model construction and evaluation

The classification performance of different modelling methods is illustrated in **Table 4**. The inputs to the model were 2 clinical features and 10 radiomics features, as shown in **Table 3**.

On the original imbalanced dataset, the SVM has a severe category bias, testing with a SPEC of 1.00 but a SENS of only 0.36. The cost-sensitive SVM model obtains a small increase in SENS of 0.36 to 0.45 but is accompanied by a large decrease in SPEC of 1.00 to 0.84. The AdaBoost shows a better performance than the cost-sensitive SVM, while SPEC decreased by only 0.09.

Training the SVM and AdaBoost models on the balanced dataset resulted in a substantial increase in SENS compared to the results from the original imbalanced dataset, both reaching 0.73, but correspondingly, a substantial decrease in SPEC, from 1.00 to 0.78 and from 0.91 to 0.75, respectively.

The SMOTEBoost's SENS is equal to that of the AdaBoost trained on the original imbalanced dataset, but its SPEC is only 0.88, lower than AdaBoost's 0.91. RUSBoost shows better classification performance than other methods, with a SENS of 0.72, a SPEC of 0.82, an ACC of 0.82, and an AUC of 0.83 by bootstrap validation in the training set. Moreover, the model has excellent classification performance with a SENS of 0.82, a SPEC of 0.84, an ACC of 0.84, and an AUC of 0.87, in the test set. The confusion matrix and ROC curves are shown in **Figure 4**.

3.4 The effect of our feature set

The verification result of feature set effectiveness is illustrated in **Table 5**. The model built with the feature of GA alone has a high SPEC of 0.88 and low SENS of 0.58. For the combination of GA and GDM, there is an increase in SENS of 0.58 to 0.72, but SPEC decreases by 0.19. The best classification performance can be achieved only with our designed feature set that includes radiomics features, GA, and GDM together.

As a validation measure of the stability of the feature set, each image was additionally calibrated with a square ROI. The square ROI was outlined within the fetal lung region as shown in **Figure 5**. Feature extraction was performed from the square ROI and the manually delineated ROI separately. As illustrated

in **Table 5**, the square ROI and the manual ROI achieved similar performance outcomes. There was only a difference of 0.01 in the ACC and 0.02 in the AUC and SENS, with the same specificity.

4 Discussion

Prenatal prediction and therapy for NRM are an effective way to improve the quality of life of NRM newborns. There is a consensus to study non-invasive methods to predict NRM using fetal lung ultrasound images. However, there is no unified feature set for the prenatal prediction of NRM, and the dataset collected in medical practice is often imbalanced and few-shot. To tackle these challenges, our study focuses on the design of feature sets with a strong representation of fetal lung ultrasound images and effective classification modelling methods.

4.1 The feature set for predicting NRM

Considering that the fetal lung in the ultrasound image is homogeneous, we designed radiomics features based on the image greyscale and texture, which can avoid the influence of the ROI's size and location on feature extraction. For each fetus, 380 radiomics features were extracted from the fetal lung region of ultrasound images, and 10 of them were selected for modelling. The energy of horizontal, which characterizes the brightness in the horizontal direction of the wavelet transform, has a mean value of 1400 in normal fetal lungs, which is higher than 1200 in NRM fetal lungs. The high grey-level run emphasis of the normal fetal lung has a higher mean value of 298 than the NRM fetal lungs of 279, which means that the fetal lung region is more homogeneous in normal fetal lungs than NRM fetal lungs. For the long-run high grey-level emphasis of vertical feature, the mean value of the normal fetal lungs is 432, which is smaller than that of the NRM fetal lung of 462, which suggests that the fetal lung region is more delicate in normal fetal lungs than NRM fetal lungs. It can be concluded that the lung region of normal fetuses has a more delicate and homogeneous texture on the ultrasound image and is brighter than that of NRM fetuses. The features we selected were also stable. The radiomics features extracted from the square ROI and the manual ROI achieved similar performance outcomes with the same modelling method (the difference was less than 0.2 for each measure), as shown in **Table 5**.

In addition to radiomics features, GA and GDM, two clinical features identified to be strongly correlated with NRM, were also added to the feature set. Newborns with a low GA have a significantly increased risk of NRM due to immature lungs, and GDM in pregnant women leads to delayed lung development in the fetus, increasing the risk of NRM. As shown in **Table 5**, the combination of GDM and GA obtained an increase from 0.8 to 0.83 in the AUC and from 0.58 to 0.72 in SENS compared to the prediction by only GA. With the addition of radiomics features, the SPEC and SENS were both significantly improved. In conclusion, the feature set designed in this study that includes radiomics features, GA, and GDM is more effective for NRM prediction and is not affected by the size or location of the ROI.

4.2 Model development

Imbalance and few-shot are inevitable in medical datasets, which pose many challenges for modelling. As shown in **Table 4**, there is a large class bias and poor classification performance on small imbalanced datasets using the conventional SVM. The methods of data augmentation, cost-sensitive learning, and ensemble learning are commonly used on imbalanced few-shot datasets. Here, these methods were performed and analysed to find the most effective modelling method.

The cost-sensitive SVM and AdaBoost show an improvement of 0.21 and 0.36 in SENS compared with the SVM in **Table 4**, but there is a decrease of 0.10 and 0.15 in SPEC in the training set. As for the cost-sensitive SVM, since there are few NRM samples, a higher cost is needed, which makes the compression of boundaries more severe, and the classifier tends to sacrifice multiple normal samples to ensure that one NRM sample is correct with a sharp decline in the generalization performance. The AdaBoost has a better performance than cost-sensitive SVM, with a SENS of 0.68 and a SPEC of 0.84. The ensemble learning method's lower overfitting allows it to exhibit a better generalization performance than the individual learner SVM or the cost-sensitive SVM.

Training on the balanced training set augmented with ADASYN, the SVM and AdaBoost does not show a significant improvement compared to training on the original imbalanced dataset, with an increase of 0.35 and 0.23 in SENS and a decrease of 0.25 and 0.26 in SPEC. For better illustration, we used t-SNE [22] to visualize the sample distribution of the original dataset and the balanced dataset augmented by ADASYN. As shown in **Figure 6**, there is aliasing between normal and NRM samples, making it difficult to classify. By generating pseudo-samples around the minority class, ADASYN leads the classifier to draw more attention to the NRM samples. However, it also exacerbates aliasing and results in poor classification performance. The generated pseudo-samples also tend to introduce plenty of noise, especially when the aliasing of samples is terrible. The data augmentation method is not appropriate in our application.

The SENS of SMOTEBoost is still low because aliasing in the dataset makes SMOTE introducing considerable noise. RUSBoost shows better classification performance than other methods. It reaches a SENS of 0.72, a SPEC of 0.82, an ACC of 0.82, and an AUC of 0.83 in the training set and a SENS of 0.82, a SPEC of 0.84, ACC of 0.84, and an AUC of 0.87 in the test set. RUSBoost can reduce overfitting and improve the classification model's generalization ability by combining weak base learners and bootstrap sampling with the AdaBoost algorithm. The input dataset of each learner is obtained by bootstrap undersampling, which enriches the sample distribution that the base learners have learned and reduce the effects of imbalance. The drawback of massive sample loss of undersampling in a small dataset is compensated by ensemble learning, while random undersampling ensures that the samples are real and avoids the noise that caused by data augmentation.

4.3 Strengths and limitations

Our study has three strengths. First, to the best of our knowledge, this is the first study to incorporate GDM, GA, and radiomics features for NRM prenatal prediction. The diagnostic efficacy of the model we

developed based on fetal lung ultrasound images in this study reached which are similar to those of many previous reports of amniocentesis [23] [24] [25]. Second, we developed a practical modelling approach to address the problems of imbalance and few-shot. RUSBoost shows excellent performance and generalization capabilities compared with the other methods used for comparison in this study. Third, we used radiomics features based on the image greyscale and texture for the prenatal prediction of NRM, whose performance is efficient and robust, without the influences of segmentation results.

As a retrospective study, this study has some limitations that should be acknowledged. Clinical outcome of the fetuses depends on several clinical factors. In addition to GA and GDM, more clinical information could be studied for its correlation with fetal lung development and used for NRM prediction. A comparative study on the right and left lungs to verify the generalizability of the method between the right and left lungs is also needed. As for those limitations, a multicentre study is underway.

5 Conclusion

In conclusion, our results show that the radiomics features of the fetal lung can be used as an efficient and robust biomarker for NRM prediction. The diagnostic efficacy of the model based on fetal lung ultrasound images, which incorporates routinely available clinical characteristics GA and GDM and radiomics features, achieves a better clinical outcome, which might afford a non-invasive tool that is easy to implement in NRM prediction.

Declarations

Funding. This research was supported by the National Natural Science Foundation of China (Grants 61871135, 81627804 and 81830058) and the Science and Technology Commission of Shanghai Municipality (Grants 20DZ1100104).

Conflicts of interest. The authors declare that they have no conflict of interest.

Availability of data and material. The processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

Code availability. Access via a formal application process, specifying the conditions that an applicant may need to meet (nationality, membership in professional associations, security clearance, etc.).

Authors' contributions. All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Jing Jiao and Yanran Du. The first draft of the manuscript was written by Jing Jiao and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Ethics approval. This study was approved by the Ethics Committee of Obstetrics and Gynecology Hospital Affiliated to Fudan University, Shanghai, China.

Consent to participate. Informed consent was obtained from all individual participants included in the study.

Consent for publication. Patients signed informed consent regarding publishing their data and photographs.

References

1. Teune M, Bakhuizen S, Bannerman C, et al. A systematic review of severe morbidity in infants born late preterm[J]. American Journal of Obstetrics and Gynecology, 2011, 205(4): 374.e1-374.e9.
2. Clark S, Miller D, Belfort M, et al. Neonatal and maternal outcomes associated with elective term delivery[J]. Am J Obstet Gynecol. 2009;200(2):156. .e1-156.e4..
3. Yarbrough M, Grenache D, Gronowski A. Fetal lung maturity testing: the end of an era[J]. Biomark Med. 2014;8(4):509–15.
4. Jobe A, Goldenberg R. Antenatal corticosteroids: an assessment of anticipated benefits and potential risks[J]. Am J Obstet Gynecol. 2018;219(1):62–74.
5. Palacio M, Bonet-Carne E, Cobo T, et al. Prediction of neonatal respiratory morbidity by quantitative ultrasound lung texture analysis: a multicenter study[J]. American journal of obstetrics gynecology. 2017;217(2):196. e1-196. e14.
6. Bonet-Carne E, Palacio M, Cobo T, et al. Quantitative ultrasound texture analysis of fetal lungs to predict neonatal respiratory morbidity[J]. Ultrasound in Obstetrics Gynecology. 2015;45(4):427–33.
7. Azad M, Moyce B, Guillemette L, et al. Diabetes in pregnancy and lung health in offspring: developmental origins of respiratory disease[J]. Paediatr Respir Rev. 2017;21:19–26.
8. Winn H, Klosterman A, Amon E, et al. Does preeclampsia influence fetal lung maturity[J]. J Perinat Med. 2000;28(3):210–3.
9. Seiffert C, Khoshgoftaar T, Van Hulse J, et al. RUSBoost: a hybrid approach to alleviating class imbalance[J]. IEEE Transactions on Systems Man Cybernetics-Part A: Systems Humans. 2009;40(1):185–97.
10. Chawla N, Lazarevic A, Hall L, et al. SMOTEBoost: improving prediction of the minority class in boosting[C]. European Conference on Principles of Data Mining and Knowledge Discovery. Springer, Berlin, Heidelberg, 2003: 107–119.
11. Aerts H, Velazquez E, Leijenaar R, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach[J]. Nat Commun. 2014;5(1):1–9.
12. Han S, Lee H, Choi J. Computer-aided prostate cancer detection using texture features and clinical features in ultrasound image[J]. J Digit Imaging. 2008;21(1):121–33.

13. Haralick R, Shanmugam K, Dinstein I. Textural features for image classification[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1973 (6): 610–621.
14. Chu A, Sehgal C, Greenleaf J. Use of gray value distribution of run lengths for texture analysis[J]. Pattern Recogn Lett. 1990;11(6):415–9.
15. Galloway M. Texture analysis using grey level run lengths[J]. Computer Graphics Image Processing. 1974;75:172–99.
16. Thibault G, Fertil B, Navarro C, et al. Shape and texture indexes application to cell nuclei classification[J]. Int J Pattern Recognit Artif Intell. 2013;27(01):1357002.
17. Amadasun M, King R. Textural features corresponding to textural properties[J]. IEEE Transactions on Systems Man Cybernetics. 1989;19(5):1264–74.
18. Loh WY. Regression Trees With Unbiased Variable Selection and Interaction Detection[J]. Stata Sinica. 2002;12(2):361–86.
19. He H, Bai Y, Garcia E. ADASYN: adaptive synthetic sampling approach for imbalanced learning[C]. IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, 2008:1322–1328.
20. Cao Q, Wang SZ. Applying Over-sampling Technique Based on Data Density and Cost-sensitive SVM to Imbalanced Learning[C]. International Conference on Information Management. IEEE, 2011.
21. Freund Y, Schapiro R. A decision-theoretic generalization of on-line learning and an application to boosting[J]. J Comput Syst Sci. 1995;55:119–39.
22. Laurens V, Hinton G. Visualizing data using t-SNE[J]. Journal of Machine Learning Research. 2008;9(Nov):2579–605.
23. Wijnberger LD, Huisjes AJ, Voorbij HA, et al. The accuracy of lamellar body count and lecithin/sphingomyelin ratio in the prediction of neonatal respiratory distress syndrome: A meta-analysis[J]. BJOG. 2001;108(6):583–8.
24. Karcher R, Sykes E, Batton D, et al. Gestational age-specific predicted risk of neonatal respiratory distress syndrome using lamellar body count and surfactant-to-albumin ratio in amniotic fluid[J]. Am J Obstet Gynecol. 2005;193(5):1680–4.
25. Haymond S, Luzzi VI, Parvin CA, et al. A direct comparison between lamellar body counts and fluorescent polarization methods for predicting respiratory distress syndrome[J]. Am J Clin Pathol. 2006;126(6):894–9.

Tables

Table 1 The summary of the feature set we designed for predicting NRM		
Feature type	Feature name	Feature number
Clinical information	(1) GA, (2) GDM	2
Greyscale histogram features	(3) Energy, (4) Entropy, (5) Kurtosis, (6) Mean, (7) Median absolute deviation, (8) Median, (9) Range, (10) Uniformity, (11) Variance, (12) Root mean square, (13) Skewness, (14) Deviation, (15) Histogram kurtosis, (16) Histogram mean, (17) Histogram variance, (18) Histogram skewness.	16
ROI textural features	(19) Mean of contrast, (20) SD of contrast, (21) Mean of covariance, (22) SD of covariance, (23) Mean of non-similarity, (24) SD of non-similarity.	6
GLCM textural features	(25) Energy, (26) Entropy, (27) Dissimilarity, (28) Contrast, (29) Inversed difference, (30) Correlation 1, (31) Correlation 2, (32) Homogeneity, (33) Autocorrelation, (34) Cluster shade, (35) Cluster prominence, (36) Maximum probability, (37) Sum of squares, (38) Sum average, (39) Sum variance, (40) Sum entropy, (41) Difference variance, (42) Difference entropy, (43) Information measures of correlation 1, (44) Information measures of correlation 2, (45) Maximal correlation coefficient, (46) Inverse difference normalized, (47) Inverse difference moment normalized.	23
GLRLM textural features	(48) Short-run emphasis, (49) Long-run emphasis, (50) Grey-level non-uniformity, (51) Run length non-uniformity, (52) Run percentage, (53) Low grey-level run emphasis, (54) High grey-level run emphasis, (55) Short-run low grey-level emphasis, (56) Short-run high grey-level emphasis, (57) Long-run low grey-level emphasis, (58) Long-run high grey-level emphasis, (59) Grey-level variance, (60) Run-length variance.	13
GLSZM textural features	(61) Small zone emphasis, (62) Large zone emphasis, (63) Grey-level non-uniformity, (64) Zone size non-uniformity, (65) Zone percentage, (66) Low grey-level zone emphasis, (67) High grey-level zone emphasis, (68) Small zone low grey-level emphasis, (69) Small zone high grey-level emphasis, (70) Large zone low grey-level emphasis, (71) Large zone high grey-level emphasis, (72) Grey-level variance, (73) Zone-size variance.	13
NGTDM textural features	(74) Coarseness, (75) Contrast, (76) Busyness, (77) Complexity, (78) Strength.	5
Wavelet features	(79-154) Approximation, (155-230) Horizontal, (231-306) Vertical, (307-382) Diagonal.	304
Total feature number		382
<p>GA: gestational age, GDM: gestational diabetes mellitus, ROI: region of interest (fetal lung region), SD: standard deviation,</p> <p>GLCM: grey-level co-occurrence matrix, GLRLM: grey-level run-length matrix, GLSZM: grey-level size zone matrix,</p> <p>NGTDM: neighbourhood grey-tone difference matrix.</p>		

Approximate, horizontal, vertical, and diagonal were decomposed from the image by wavelet transform (first-level decomposition).

1. Clinical information: GA and GDM are strongly correlated with NRM[7][8]. GA was determined by the last menstrual period and verified by first-trimester dating ultrasound (crown-rump length). According to the presence of GDM during pregnancy, these pregnant women were divided into Yes and No groups.
2. Greyscale histogram features: Describe the greyscale and histogram distribution of the ROI in fetal lung ultrasound images[11].
3. Textural features: Describe detailed, invisible greyscale changes and associations in fetal lung ultrasound images.
 1. ROI textural features: Describe the distribution of greyscale inside the ROI[12].
 2. Grey-level co-occurrence matrix (GLCM) textural features: Describe the specified spatial linear relationship between the frequencies of two greyscale intensities inside the ROI[13].
 3. Grey-level run-length matrix (GLRLM) textural features: Describe the roughness of the texture by calculating the run-length of the collinear image pixels of the same grey-level in a given direction inside the ROI[14][15].
 4. Grey-level size zone matrix (GLSZM) textural features: Describe the uniformity of the small pixel population of the ROI[13][16].
 5. Neighbourhood grey-tone difference matrix (NGTDM) textural features: Describe the difference between the greyscale of each image pixel and the greyscale of its neighbours inside the ROI[17].
4. Wavelet features: Describe information that is not directly reflected by the greyscale and textural features of the original image. Every fetal lung ultrasound image was decomposed into four components: approximate, horizontal, vertical, and diagonal by wavelet transform (first-level decomposition). Then, the 76 features mentioned above were extracted separately on each component. Finally, a total of 304 wavelet features were extracted.

Table 2 Characteristics of the training set and test set						
Characteristics	Training set (<i>n</i> = 167)			Test set (<i>n</i> = 43)		
	Normal	NRM	<i>p</i> value	Normal	NRM	<i>p</i> value
No. of images	127	40	-	32	11	
GA*	36.49	34.37		36.78	34.53	
Birth weight (g)*	3096	2978		3145		
<i>GDM</i>						
Yes	48(37.80%)	26(65.00%)	-	10(31.25%)	7(63.64%)	-
No	79(62.20%)	14(35.00%)	-	22(68.75%)	4(36.36%)	-
<i>Mode of delivery</i>			0.35	0.94		
Spontaneous vaginal delivery	56(44.09%)	21(52.50%)	-	15(46.88%)	5(45.45%)	-
Caesarean delivery	71(55.91%)	19(47.50%)	-	17(53.12%)	6(54.55%)	-
<i>Gender of newborn</i>			0.87	0.43		
Female	59(46.46%)	18(45.00%)	-	16(50.00%)	7(63.64%)	-
Male	68(53.54%)	22(55.00%)	-	16(50.00%)	4(36.36%)	-
<i>Apgar</i>			-	-		
5 min ≤ 7	0(0.00%)	4(10.00%)	-	0(0.00%)	0(0.00%)	-
5 min > 7	127(100.00%)	36(90.00%)	-	32(100.00%)	11(100.00%)	-
<i>GA</i> : gestational age, <i>GDM</i> : gestational diabetes mellitus.						
*Data are means ± standard deviations.						
The <i>t</i> test was performed for continuous variables and the X ² test was performed for categorical variables.						

Table 3 Feature names and means of the features selected		
Feature name	Mean std	
	Normal	NRM
Energy		
Inverse difference moment normalized	4	
High grey-level run emphasis		
Run-length variance		
Inverse difference moment normalized of approximation		
Information measure of correlation 1 of approximation		
Energy of horizontal	036	
Sum entropy of vertical		
Long-run high grey-level emphasis of vertical		
Energy of diagonal		
Approximate, horizontal, vertical, and diagonal were decomposed from the image by wavelet transform (first-level decomposition).		

Table 4 The classification performance of different modelling methods										
Method	Training set (mean std)				Test set					
	ACC	AUC	SENS	SPEC	ACC	AUC	SENS	SPEC		
<i>Original imbalanced training set</i>										
SVM	0.81 0.04	0.76 0.07	<i>0.32 0.11</i>	0.99 0.02	0.84	0.78	<i>0.36</i>	1.00		
AdaBoost	0.79 0.10	0.72 0.16	0.68 0.18	0.84 0.09	0.81	0.79	0.55	0.91		
Cost-sensitive SVM	0.79 0.07	0.73 0.10	0.43 0.21	0.89 0.09	<i>0.74</i>	<i>0.75</i>	0.45	0.84		
<i>Balanced training set augmented with ADASYN</i>										
SVM	0.72 0.08	0.79 0.10	0.67 0.17	<i>0.74 0.11</i>	0.77	0.85	0.73	0.78		
AdaBoost	<i>0.71 0.07</i>	0.71 0.08	0.55 0.15	0.76 0.07	<i>0.74</i>	0.82	0.73	<i>0.75</i>		
<i>Original imbalanced training set (combining data balance and ensemble learning)</i>										
SMOTEBoost	0.79 0.05	<i>0.70 0.09</i>	0.52 0.14	0.89 0.10	0.79	0.80	0.55	0.88		
RUSBoost	0.82 0.10	0.83 0.13	0.72 0.15	0.82 0.12	0.84	0.87	0.82	0.84		
The best results of each metric are shown in bold, and the worst results are shown in italics. Performance evaluation results obtained by bootstrap K-fold cross-validation in the training set.										

Table 5 The classification performance of RUSBoost with different features on the original imbalanced few-shot dataset				
Feature	Training set (mean std)			
	ACC	AUC	SENS	SPEC
GA	0.81 0.07	0.8 0.10	0.58 0.21	0.88 0.06
GA & GDM	0.71 0.12	0.83 0.10	0.72 0.24	0.69 0.15
<i>Radiomics features extracted from the ROI of the manually delineated fetal lung region</i>				
Gestational age, GDM & radiomics features	0.82 0.10	0.83 0.13	0.72 0.05	0.82 0.12
<i>Radiomics features extracted from the square ROI within the fetal lung region</i>				
Gestational age, GDM & radiomics features	0.81 0.14	0.81 0.09	0.70 0.07	0.82 0.15
The best results of each metric are shown in bold, and the worst results are shown in italics. Performance evaluation results obtained by bootstrap K-fold cross-validation in the training set.				

Figures

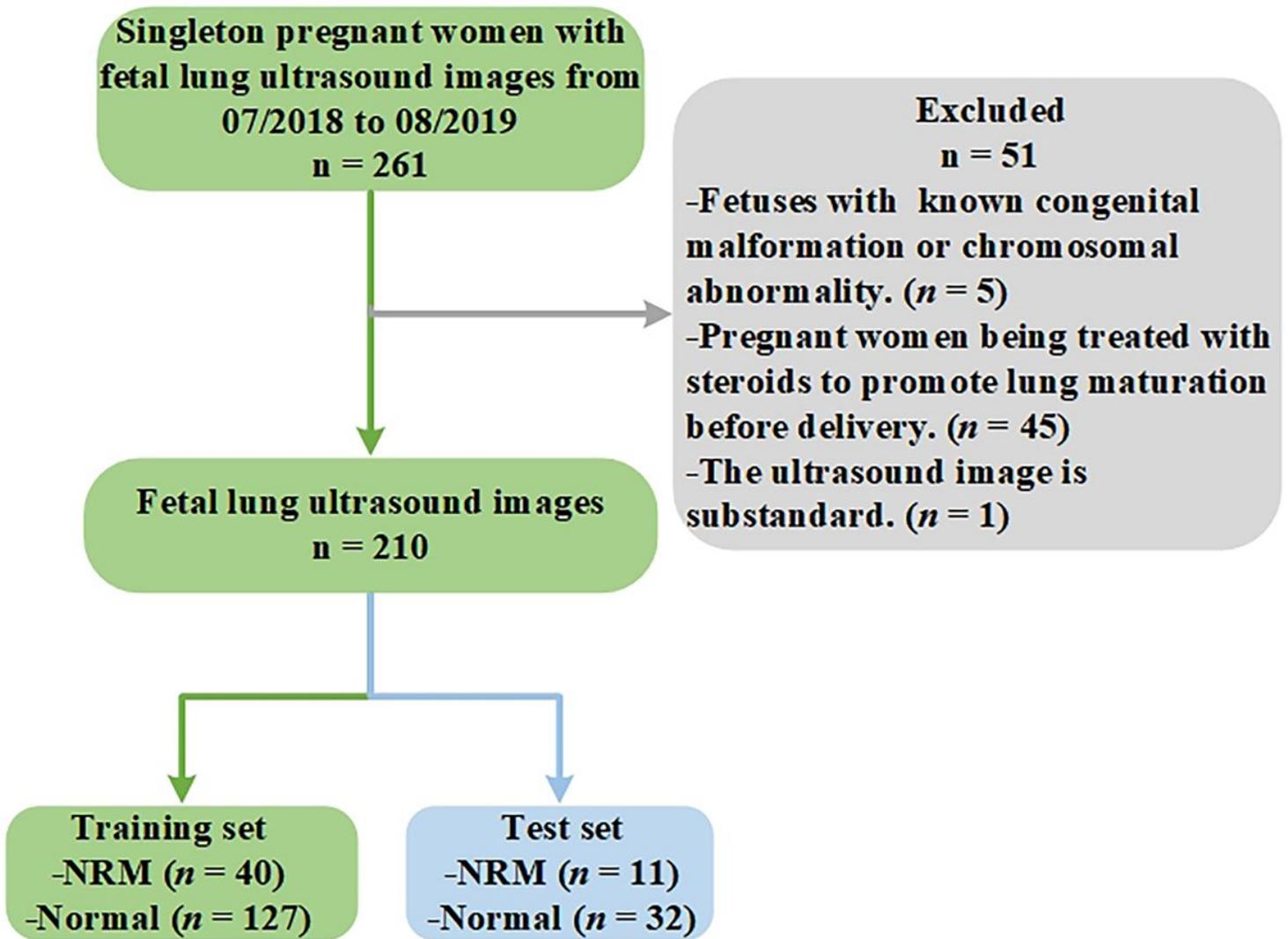


Figure 1

Flowchart of the selection process of the study population. Pregnant women who met the following criteria were enrolled in the study: 1) singleton pregnancy; 2) those with complete medical information who had undergone maternity examination and subsequent delivery in our hospital; 3) fetuses with no known congenital malformation or chromosomal abnormality; 4) those with no diabetes before pregnancy; and 5) those who had not been prescribed steroids before delivery. Finally, a total of 210 singleton pregnant women with 210 fetal lung ultrasound images were enrolled in our study and randomly divided into the training set and test set at a ratio of approximately 8:2. It is worth noting that we kept the same proportion of NRM and normal in both sets. The training set contains 167 images, of which 40 are NRM and 127 are normal. The test set contains 43 images, of which 11 are NRM and 32 are normal.

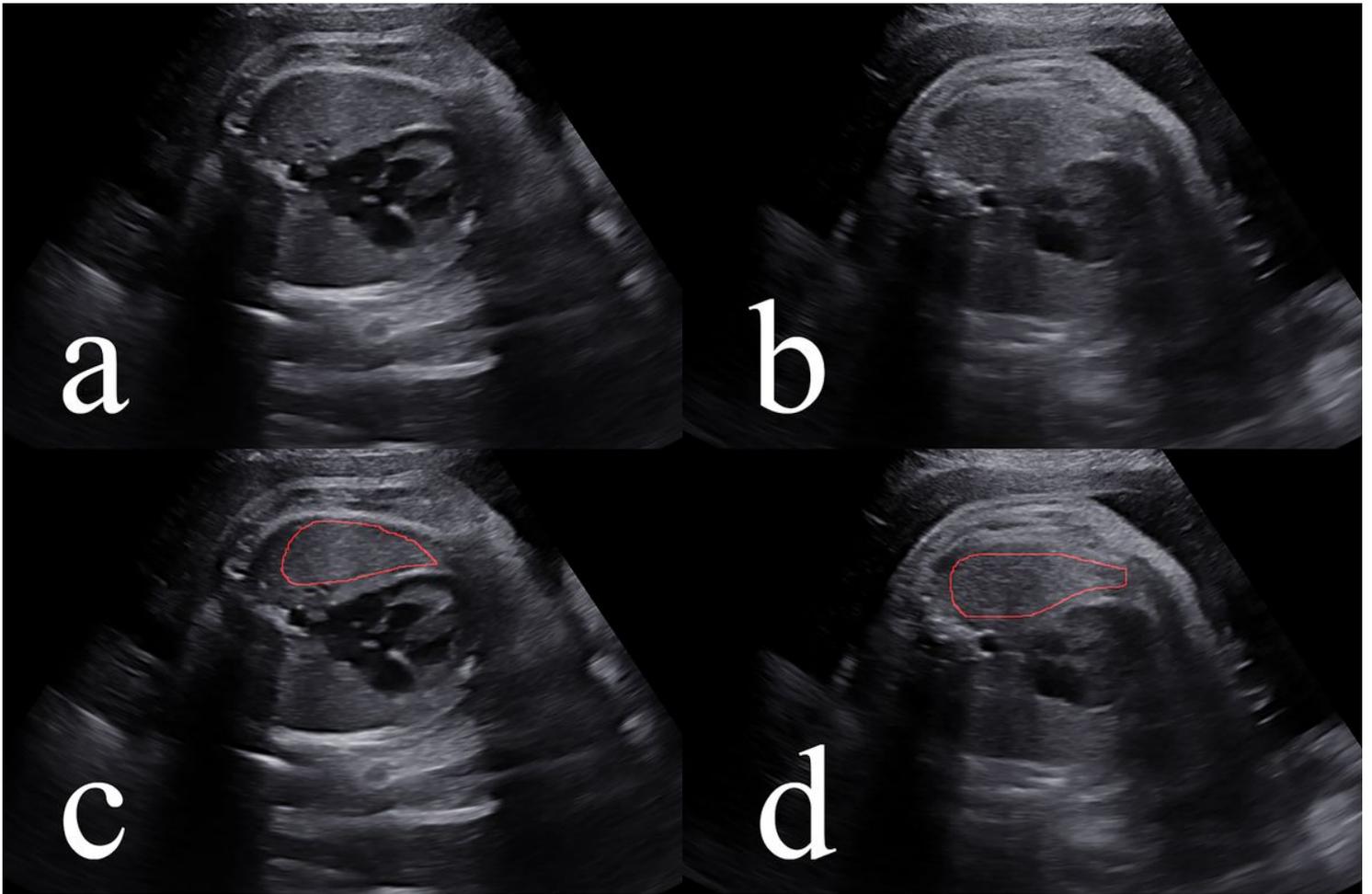


Figure 2

Examples of NRM and normal fetal lung ultrasound images. a) An NRM fetal lung ultrasound image. b) A normal fetal lung ultrasound image. c) and d) are the manual delineations of the ROIs. The manual delineation of the ROI in each fetal lung image was performed by two radiologists who were blinded to the medical histories of the pregnant women and the neonatal outcomes. Each image was delineated by one of these physicians and confirmed by the other physician.

Box plots of the top 3 features with a high correlation between the normal and NRM fetal lung ultrasound images of the selected features.

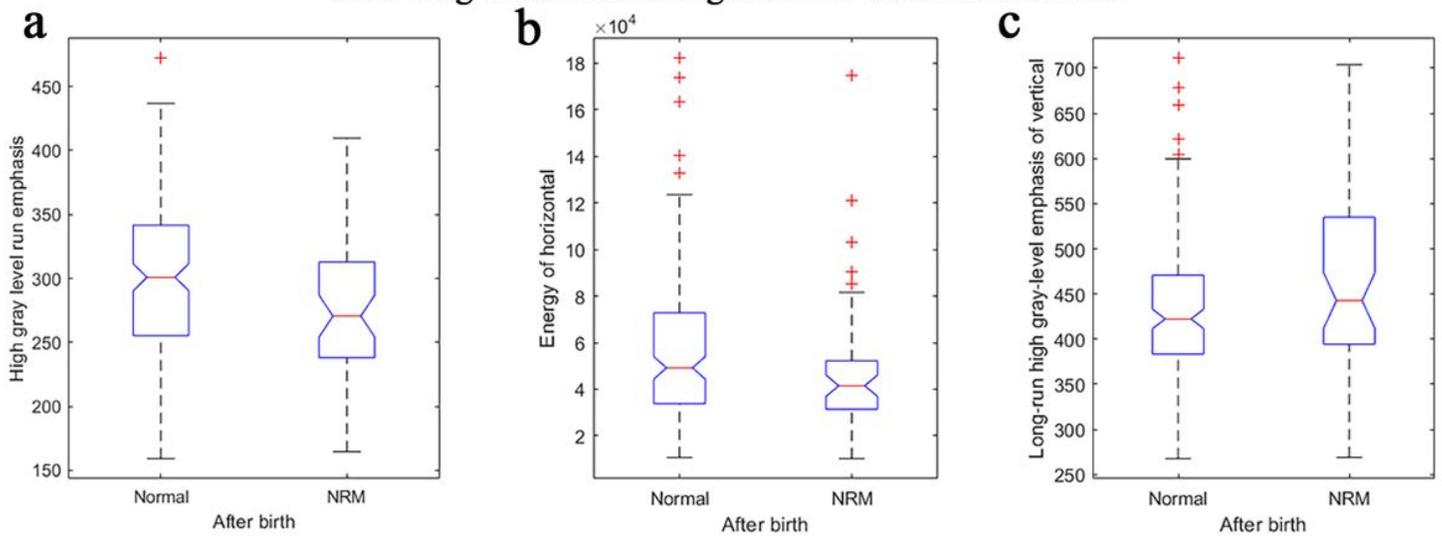


Figure 3

Box plots of the top 3 features of the 10 selected features. a), b) and c) are the box plots of the high grey-level run emphasis, energy of horizontal and long-run high grey-level emphasis of vertical features extracted from the ROIs of the normal and NRM samples. The normal fetal lung has higher mean values for the features of high grey-level run emphasis (298 ± 62.5) and energy of diagonal (1400 ± 724) than the NRM. For the long-run high grey-level emphasis of vertical feature, the mean value of the normal fetal lung is 432, which is smaller than that of the NRM of 462.

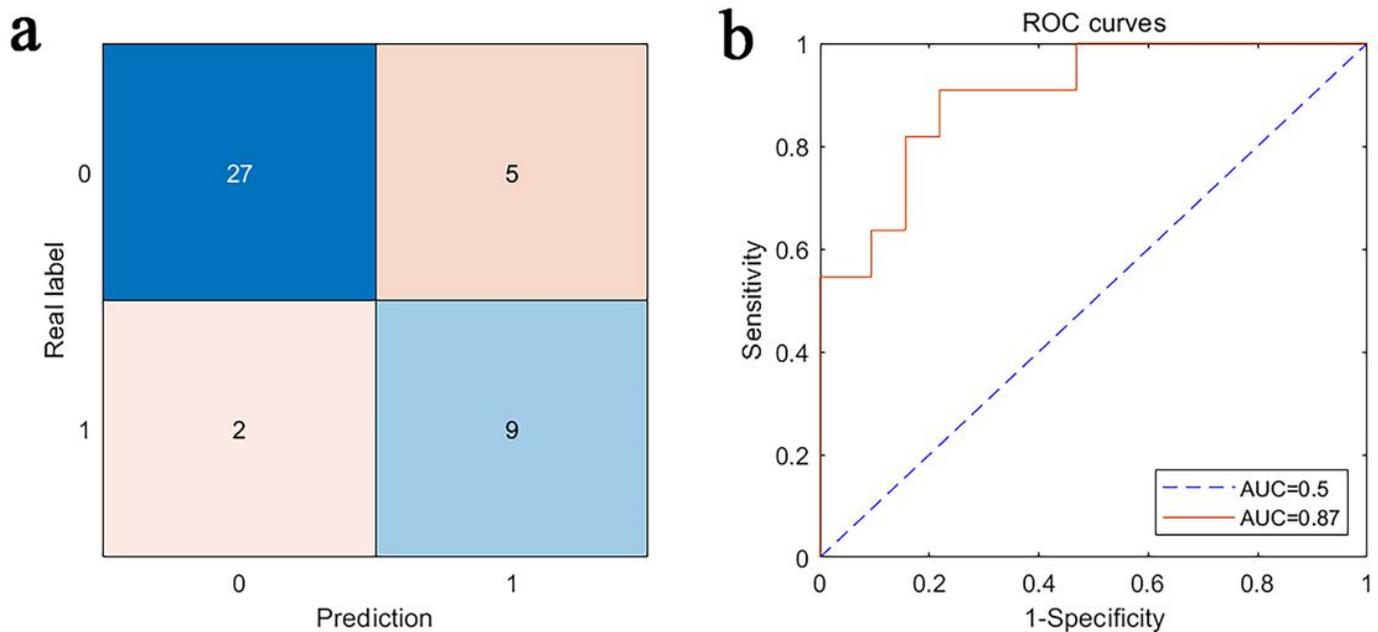


Figure 4

The confusion matrix and ROC curves tested in the test set. The confusion matrix (a) and ROC curves (b) was acquired with the final model tested in the test set. In the test set, nine NRM samples were correctly

predicted, while only two NRM samples were misclassified as normal. Only 5 of 32 normal samples were misclassified as NRM, with a low false positive rate.

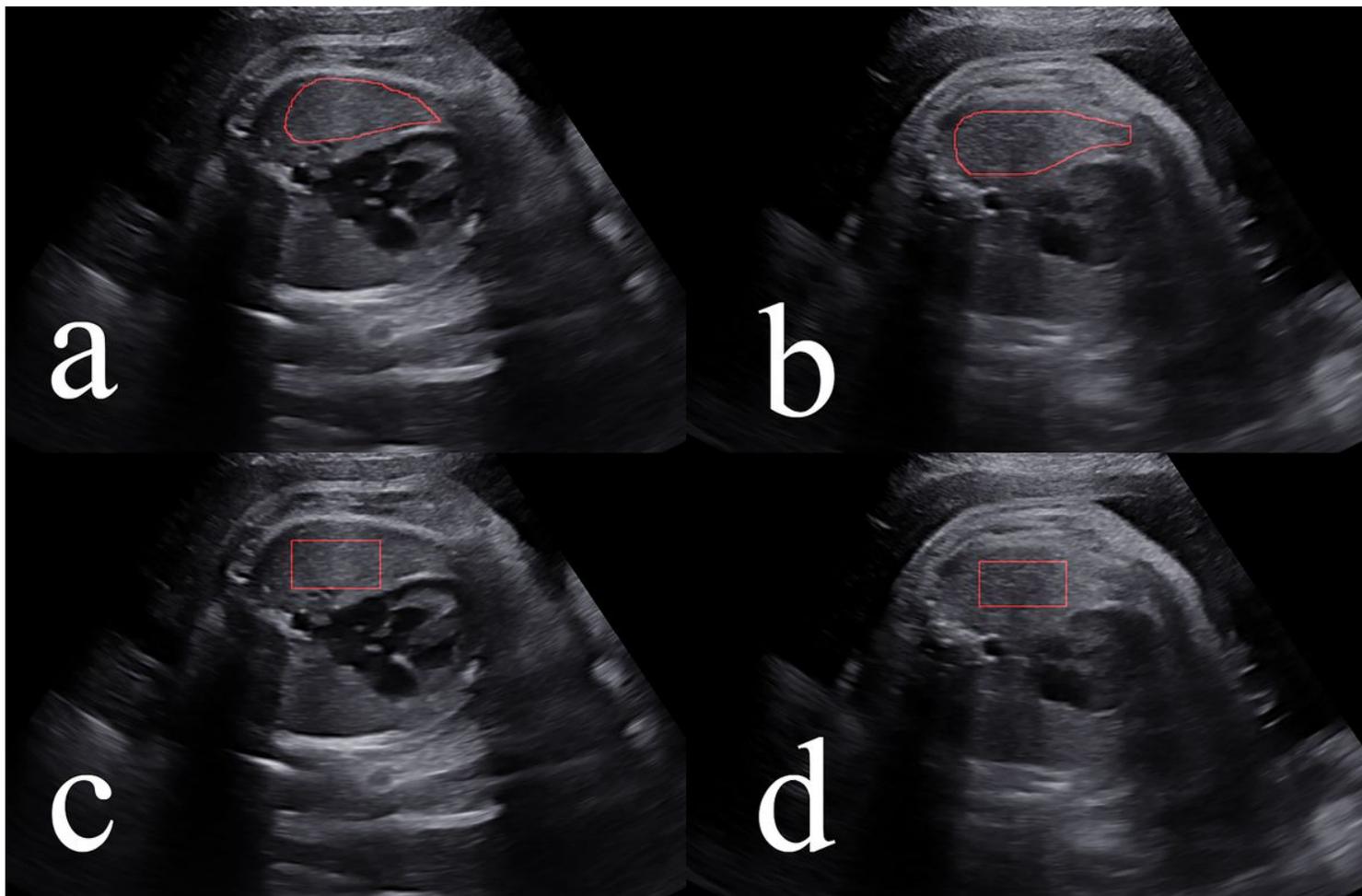


Figure 5

Examples of different ROIs in the NRM and normal fetal lung ultrasound images. a) and b) are manually delineated ROIs within the fetal lung region in an NRM and normal fetal lung ultrasound image. c) and d) are square ROIs within the fetal lung region in an NRM and normal fetal lung ultrasound image.

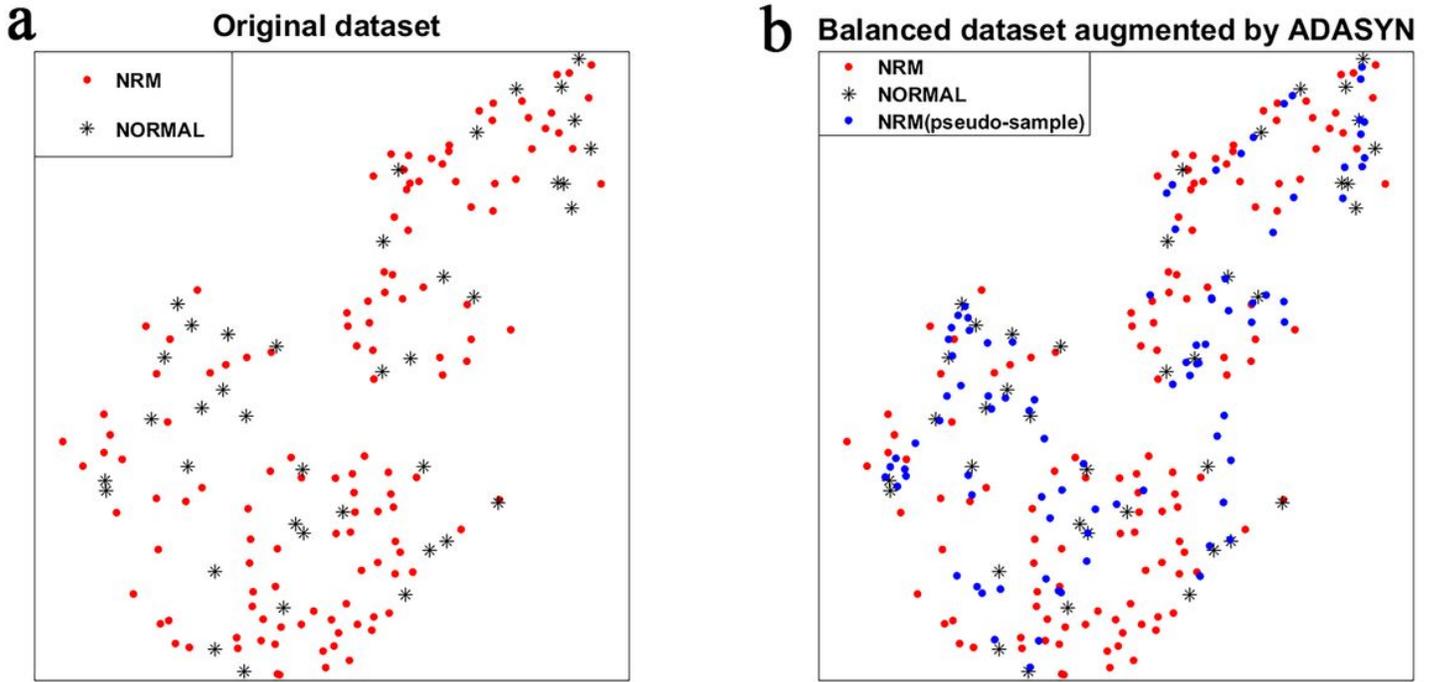


Figure 6

The distribution of the samples. a) The sample distribution of the original dataset with terrible category aliasing. b) The sample distribution of the balanced dataset augmented by ADASYN.