





Fetal Weight Percentile Classification Across Gestational Weeks by Comparing Machine-Learning Algorithms Using Ultrasound Images

Maryam Mehdipor Ghazvini¹, Vahab Dehlaghi², Nazanin Farshchian³ , Hamid Sharini^{2,*} 

¹ Department of Biomedical Engineering, School of Medicine, Kermanshah University of Medical Sciences, Kermanshah, Iran

² Department of Biomedical Engineering, Faculty of Medicine, Kermanshah University of Medical Sciences, Kermanshah, Iran

³ Department of Radiology, Imam Reza Hospital, School of Medicine Taleghani Hospital, Kermanshah University of Medical Sciences, Kermanshah, Iran

*Corresponding Author: Department of Biomedical Engineering, Faculty of Medicine, Kermanshah University of Medical Sciences, Kermanshah, Iran. Email: hamidsharini@gmail.com

Received: 11 April, 2026; Revised: 2 June, 2026; Accepted: 11 June, 2026

Abstract

Background: Accurate assessment of the fetal weight percentile is an important component of prenatal care, as it enables clinicians to monitor fetal growth and identify fetuses at risk of growth restriction or macrosomia. However, conventional ultrasound-based assessment may be affected by measurement variability and operator dependence. Radiomics-based machine-learning approaches may provide a more objective and reproducible framework for classifying fetal weight percentiles.

Objectives: This study aimed to compare the performance of several machine-learning algorithms for classifying fetal weight percentile categories using ultrasound images.

Methods: This analytical retrospective study was conducted at Kermanshah University of Medical Sciences using archived ultrasound data from 200 pregnant women collected during 1401 - 1402. Ultrasound images were preprocessed and denoised in MATLAB using four filters: Butterworth, Ideal, Median, and Wavelet. Regions of interest corresponding to the fetal head, abdomen, and femur were identified, and 1715 radiomics features per case were extracted using 3D Slicer. To reduce dimensionality and improve model robustness, principal component analysis was applied before classification. Machine-learning models were developed in the MATLAB Classification Learner Toolbox, including an ensemble model, a support vector machine, k-nearest neighbors, and an artificial neural network. Model performance was evaluated using 10-fold cross-validation. Evaluation metrics included accuracy, precision, sensitivity, specificity, F1-score, and the area under the receiver operating characteristic curve.

Results: Among the evaluated models, the ensemble model demonstrated the highest internal performance, with an accuracy of 0.95 and an area under the receiver operating characteristic curve of 0.93. The accuracies of the k-nearest neighbors, support vector machine, and neural network models were 0.93, 0.92, and 0.91, respectively, with corresponding area under the receiver operating characteristic curve values of 0.92, 0.91, and 0.90, respectively.

Conclusions: These findings indicate that radiomics-based machine-learning models, particularly the ensemble model, show promising performance for classifying fetal weight percentile categories from ultrasound images. Nevertheless, because this study was retrospective and relied solely on internal validation, further prospective and external validation is required before clinical application. Accordingly, the proposed model should be considered a decision-support tool rather than a replacement for clinical judgment.

Keywords: Ultrasound, Artificial Intelligence, Machine Learning, Fetal Weight Percentile

1. Background

Monitoring fetal growth in obstetrics relies heavily on estimated fetal weight (EFW), as few biomarkers

reliably predict macrosomia or fetal growth restriction. Accurate prenatal weight estimation, often expressed as a percentile for gestational age, is essential for clinical decision-making, including fetal growth surveillance,

Copyright © 2026, Journal of Clinical Research in Paramedical Sciences. This open-access article is available under the Creative Commons Attribution-NonCommercial 4.0 (CC BY-NC 4.0) International License (<https://creativecommons.org/licenses/by-nc/4.0/>), which allows for the copying and redistribution of the material only for noncommercial purposes, provided that the original work is properly cited.

How to Cite: Mehdipor Ghazvini M, Dehlaghi V, Farshchian N, Sharini H. Fetal Weight Percentile Classification Across Gestational Weeks by Comparing Machine-Learning Algorithms Using Ultrasound Images. J Clin Res Paramed Sci. 2026;15(1):e171271. doi: <https://doi.org/10.5812/jcrps-171271>

planning the mode and timing of delivery, and reducing adverse maternal and neonatal outcomes. In routine clinical practice, obstetric ultrasound provides standard biometric measurements, such as head circumference, abdominal circumference, and femur length, which are widely used for EFW calculation and fetal growth assessment (1-4). Nevertheless, ultrasound-based assessment can be affected by operator dependence, image quality, and measurement variability.

In recent years, quantitative medical image analysis has enabled the characterization of imaging data beyond visual interpretation. Radiomics refers to the extraction of a large number of quantitative features, such as intensity- and texture-related descriptors, which may capture subtle patterns not readily recognized during routine ultrasound review. Because ultrasound is widely available and routinely performed during prenatal care, integrating radiomics with machine-learning methods may improve the consistency and reproducibility of fetal growth assessment and support fetal weight percentile categorization (5).

Several studies have investigated machine-learning approaches for fetal weight prediction and classification. For example, Khan et al. evaluated classification models to categorize newborns into weight groups, including low birth weight and macrosomia, and reported strong performance for a random forest approach (6). Building on this work, the present study evaluates and compares multiple machine-learning classifiers, including an ensemble model, k-nearest neighbors (KNN), a support vector machine (SVM), and an artificial neural network (NN), for classifying fetal weight percentile categories using radiomics features extracted from fetal ultrasound images.

2. Objectives

This study aimed to develop and internally validate an automated framework integrating ultrasound imaging, radiomics feature extraction, and machine-learning classification for fetal weight percentile categorization. Given the retrospective, single-center design and the internal validation setting, the proposed approach is presented as a decision-support tool. Therefore, prospective multicenter studies and external validation are required before clinical implementation. The main contribution of this work is the development of an ultrasound-based radiomics pipeline implemented in 3D Slicer and coupled with machine-learning models to enable automated classification of fetal weight percentile categories.

3. Methods

3.1. Work Implementation Stages

The steps of the current research are shown in [Figure 1](#).

3.2. Data Acquisition

This retrospective study was conducted using archived fetal ultrasound data collected from medical training centers affiliated with Kermanshah University of Medical Sciences during 1401 - 1402, after approval by the institutional Ethics Committee. Initially, approximately 250 ultrasound cases were reviewed. Fifty cases were excluded because of inadequate image quality and/or insufficient segmentation quality that could adversely affect reliable feature extraction. Finally, 200 fetal ultrasound cases were included in the analysis.

Eligible cases were obtained from routine ultrasound examinations performed in the second and third trimesters. To meet ethical requirements, all records were de-identified before analysis; patient identifiers were removed, and all data were coded and stored confidentially.

3.3. Preprocessing

Preprocessing is a key step in ultrasound image analysis because it can enhance image quality and reduce speckle noise and artifacts, thereby enabling more reliable segmentation and feature extraction. In this study, ultrasound images were denoised and enhanced in MATLAB using 4 commonly used filters, and their outputs were compared: median, Ideal, Butterworth, and Wavelet filters.

The median filter is a nonlinear spatial filter that reduces noise by replacing each pixel intensity with the median value in a local neighborhood. The Ideal filter is a frequency-domain filter that passes a selected frequency band while suppressing frequency components outside the band. The Butterworth filter is a frequency-domain filter with a smooth transition between the passband and stopband, enabling controlled attenuation of undesired frequency components. The Wavelet filter is a multiresolution denoising approach that decomposes the image into multiple scales to suppress noise while preserving anatomical boundaries and structural details.

These filters were applied to improve the robustness of subsequent manual segmentation and radiomics feature extraction.

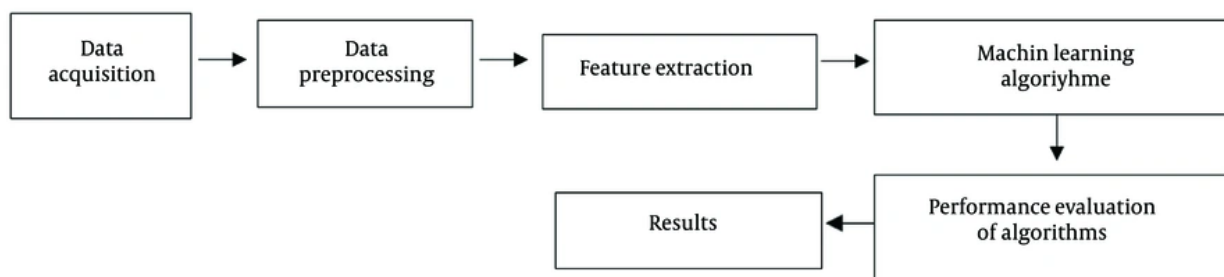


Figure 1. Workflow of the study steps

3.4. Segmentation and Radiomics Feature Extraction

After preprocessing, ultrasound images were imported into 3D Slicer version 5.2.2 for manual delineation of the regions of interest (ROIs). Segmentation was performed manually by the researcher under the supervision of an obstetrics specialist. For each case, 3 ROIs were defined, corresponding to the fetal head, fetal abdomen, and femur.

Subsequently, radiomics feature extraction was performed using the radiomics module implemented in 3D Slicer version 5.2.2. Radiomics enables the conversion of medical images into high-dimensional quantitative data by extracting features that describe the image intensity distribution, morphological characteristics, and textural heterogeneity. In the present study, multiple feature groups were extracted from each ROI, including first-order statistics, shape features, texture features, and filter-based features.

The extracted features from the 3 anatomical regions were aggregated to form a single radiomics profile for each case. Overall, 1715 radiomics features were obtained per case and used for downstream dimensionality reduction and machine-learning model development. The segmentation and feature extraction process is presented in [Figure 2](#).

3.5. Feature Matrix Construction and Outcome Labeling

Radiomics features extracted from the segmented ultrasound images were arranged into a 2-dimensional feature matrix, in which each row represented 1 case and each column represented 1 radiomics feature. The final dataset therefore consisted of 200 cases and 1715 features per case, resulting in a 200×1715 input matrix.

Fetal weight was categorized into 5 percentile-based classes according to the reference tables used in this study (7, 11): 10th - 25th percentile (class 0), 25th - 50th percentile (class 1), 50th - 75th percentile (class 2), 75th - 90th percentile (class 3), and 90th - 95th percentile (class 4). These labels were used as the target output for classifier development.

3.6. Artificial Intelligence and Machine-Learning Models

In this study, 4 supervised machine-learning classifiers were implemented and compared: ensemble, KNN, SVM, and NN. All models were trained and evaluated using the same dataset and validation framework.

3.7. Dimensionality Reduction

To reduce feature dimensionality and minimize overfitting, principal component analysis (PCA) was applied before classification. The transformed lower-dimensional feature set was then used as input for model training and testing. To prevent data leakage during model training and evaluation, PCA for dimensionality reduction was performed within each fold of the 10-fold cross-validation loop. Specifically, PCA was applied exclusively to the training data of each fold, and the resulting transformation, namely the principal components, was then used to project both the training and testing data of that fold. This procedure ensured that information from the test set did not inadvertently influence the dimensionality reduction process, thereby providing a more unbiased estimate of the model's generalization performance.

3.8. Model Validation and Performance Evaluation

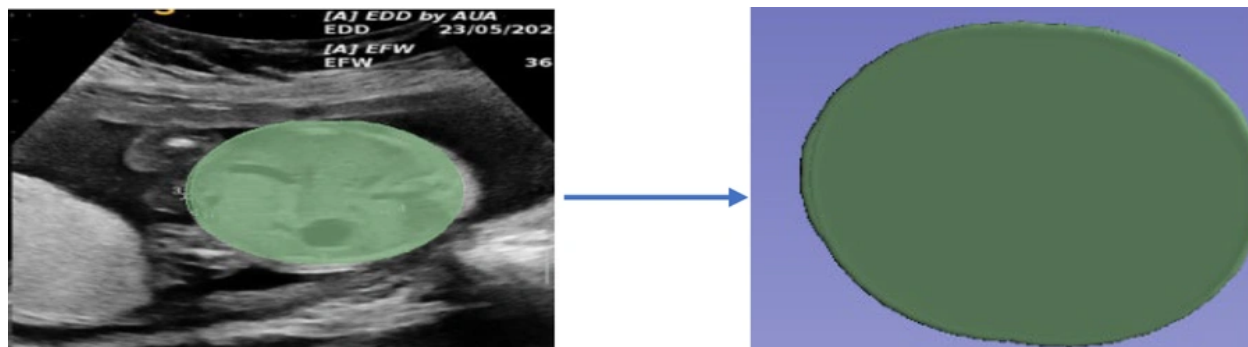


Figure 2. Segmentation of ultrasound images, including the head, abdomen, and femur, using the radiomics toolbox

Model performance was assessed using random 10-fold cross-validation. The dataset was randomly divided into 10 approximately equal folds. In each iteration, 1 fold was used for testing and the remaining 9 folds were used for training. This procedure was repeated until each fold had served once as the test set, and final performance was obtained by averaging results across folds.

The primary evaluation metric was overall accuracy, calculated from the multiclass confusion matrix as follows:

$$\text{Accuracy} = \frac{\sum k_i = 1CM_{ii}}{\sum k_{i=1} \sum k_{i=1} CM_{ij}}$$

where C is the number of classes. To provide a more comprehensive evaluation, macro-F1, balanced accuracy, and class-wise sensitivity and specificity in a one-vs-rest setting may also be reported.

4. Results

The quantitative and qualitative effects of the evaluated preprocessing filters on fetal ultrasound images are presented in [Figures 3](#) and [4](#). For the median filter, window sizes of 3, 5, 8, and 10 were tested. For the Ideal and Butterworth filters, cutoff frequencies of 10, 20, 30, and 40 were evaluated. For the Wavelet filter, decomposition/frequency levels of 1, 2, 3, and 4 were examined.

Filter performance was assessed using the signal-to-noise ratio (SNR) and mean squared error (MSE). A higher SNR indicates better preservation of useful image information relative to noise, whereas a lower

MSE indicates reduced reconstruction error and improved denoising quality.

Overall, the results indicated that increasing the median filter window size degraded performance, suggesting that smaller windows better preserve relevant anatomical boundaries. For the Ideal and Butterworth filters, increasing the cutoff frequency led to a gradual increase in MSE, and both filters exhibited broadly similar trends. For the Wavelet filter, the best performance was observed at levels 2 and 4, after which the trend changed approximately linearly. Based on the combined SNR/MSE comparison, the median filter with a window size of 3 and the Wavelet filter at levels 2 and 4 yielded the most favorable denoising performance and were therefore selected as the preferred preprocessing configurations for subsequent analysis.

After preprocessing, a total of 200 fetal ultrasound cases were analyzed. Manual segmentation was performed in 3D Slicer version 5.2.2 for 3 ROIs: fetal head, abdomen, and femur. Radiomics feature extraction was subsequently performed using the radiomics toolbox in 3D Slicer, and the features extracted from the 3 ROIs were aggregated. In total, 1715 radiomics features per case were obtained and used as input to the machine-learning models.

The fetal weight outcome was defined as a 5-class, percentile-based label, and 4 supervised classifiers, including Ensemble, KNN, SVM, and NN, were trained and evaluated using random 10-fold cross-validation. The comparative performance of the models is summarized in [Table 1](#). Overall, the ensemble-based classifier demonstrated the best performance among the evaluated algorithms for fetal weight percentile classification.

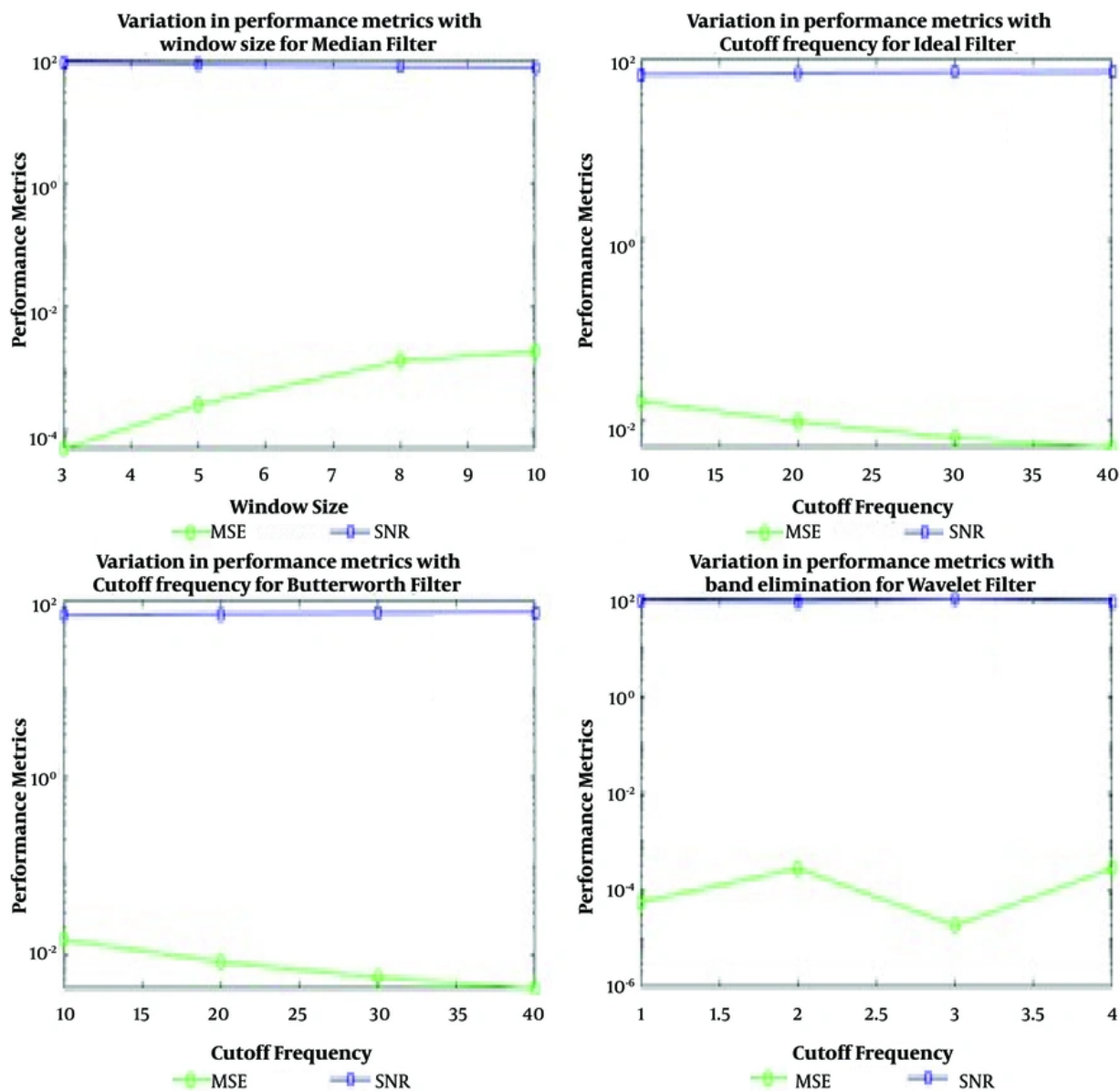


Figure 3. Performance of the filters on ultrasound images

Because this was a multiclass classification task and the class distribution may have been imbalanced, accuracy alone may not fully characterize performance. Therefore, in addition to accuracy, reporting macro-F1 and balanced accuracy is recommended to provide a more robust evaluation across all classes. The ensemble model achieved the best overall performance across all metrics, followed by KNN, SVM, and ANN.

5. Discussion

This study proposed and internally validated a radiomics-based machine-learning framework for fetal weight percentile classification using fetal ultrasound images acquired in the second and third trimesters. The workflow included image preprocessing, manual ROI segmentation of the head, abdomen, and femur,

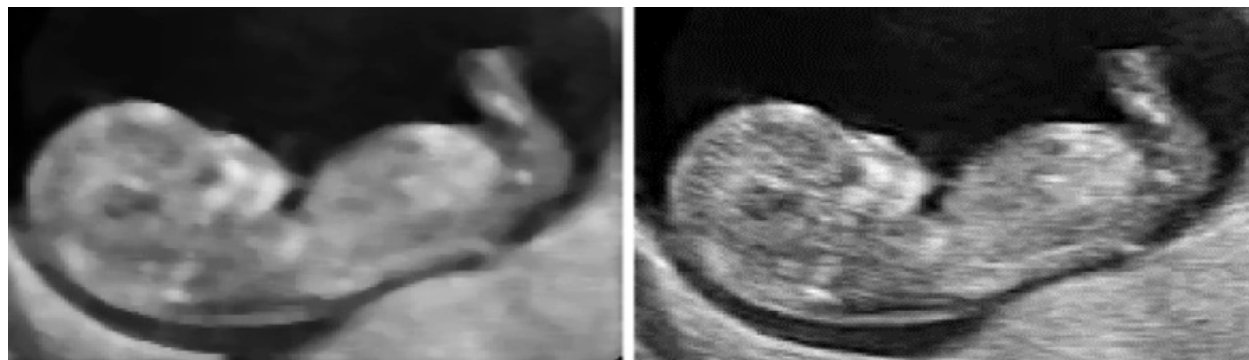


Figure 4. Noise removal from ultrasound images after applying filters

radiomics feature extraction, dimensionality reduction using PCA, and supervised classification using multiple machine-learning algorithms.

A key step in ultrasound image analysis is preprocessing because fetal ultrasound images are inherently affected by speckle noise and acquisition-related artifacts. In the present study, several commonly used filters were evaluated and compared using SNR and MSE. The findings suggested that a median filter with a small window size of 3 and wavelet-based denoising at levels 2 and 4 provided the most favorable noise reduction while preserving anatomical detail. These results are consistent with prior reports indicating that median filtering can effectively reduce noise and artifacts in medical images (8).

Following preprocessing and manual segmentation, a large number of quantitative features were extracted using radiomics. Aggregation of features across the fetal head, abdomen, and femur resulted in a high-dimensional feature space of 1715 features per case. Because high-dimensional feature sets can increase the risk of overfitting, particularly when the sample size is limited, PCA was used to reduce dimensionality before model training.

Among the evaluated classifiers, the ensemble model achieved the best overall performance for fetal weight percentile classification. This observation aligns with previous studies reporting that ensemble or hybrid strategies may improve robustness in biomedical classification tasks by combining complementary decision rules. Prior work by Tao et al. (9) and International Conference on Communications (10) similarly demonstrated that performance may vary substantially across algorithms and that combined

approaches can yield competitive results in fetal growth-related prediction tasks.

Despite these promising findings, several limitations should be acknowledged. First, this retrospective, single-center study included a moderate sample size of 200 cases, which may limit generalizability. Second, ROIs were delineated manually, which may introduce observer dependence and reduce scalability. Third, model evaluation relied on random 10-fold cross-validation without external validation; therefore, the reported performance should be interpreted as internal validation. Finally, given the multiclass setting and potential class imbalance, reporting additional metrics, such as macro-F1, balanced accuracy, and class-wise sensitivity/specificity, would provide a more complete assessment than accuracy alone.

Accordingly, the proposed approach should be considered a decision-support tool rather than a clinically deployable system. Future work should focus on external validation using independent multicenter datasets, evaluating stratified validation protocols, and investigating semiautomated or automated segmentation approaches to improve reproducibility and clinical feasibility.

5.1. Limitations and Future Directions

Several limitations of the current study warrant acknowledgment. First, the retrospective, single-center design using a dataset of 200 cases may limit the generalizability of the findings. Although rigorous random 10-fold cross-validation was employed for internal model assessment, external validation using independent datasets from diverse patient populations and clinical settings is crucial to confirm the model's

Table 1. Comparative Performance of the Evaluated Machine-Learning Models for Fetal Weight Percentile Classification^a

| Models | Accuracy | Precision | Sensitivity | Specificity | F1-Score | Geometric Mean | MCC | AUC |
|----------|----------|-----------|-------------|-------------|----------|----------------|-------|-------|
| ANN | 91.00 | 79.64 | 77.04 | 94.02 | 78.50 | 85.03 | 72.02 | 90.01 |
| SVM | 92.20 | 82.00 | 80.40 | 94.00 | 81.00 | 86.00 | 74.00 | 91.24 |
| KNN | 93.07 | 88.06 | 85.09 | 96.03 | 87.02 | 90.09 | 83.06 | 92.01 |
| Ensemble | 95.80 | 91.68 | 88.95 | 97.16 | 90.29 | 92.96 | 87.16 | 93.03 |

^a Values are expressed as percentage. Abbreviation: MCC, Matthews correlation coefficient; AUC, area under curve.

robustness and performance in real-world scenarios. Future research should prioritize the collection and analysis of such external data to further validate the proposed radiomics-based framework.

Second, manual segmentation of ROIs, although performed under expert supervision, introduces interobserver variability. Efforts to develop and validate semiautomated or fully automated segmentation algorithms could enhance reproducibility and scalability, thereby facilitating wider clinical adoption.

Third, the current model serves as a decision-support tool. Although promising, direct translation into routine clinical practice requires further prospective validation, integration into clinical workflows, and rigorous assessment of its impact on clinical decision-making and patient outcomes.

5.2. Conclusions

This study demonstrated the potential of radiomics analysis of fetal ultrasound images for fetal weight percentile classification. A radiomics-driven machine-learning pipeline applied to fetal ultrasound images classified fetal weight percentile categories with encouraging internal validation performance. Among the evaluated models, the ensemble classifier showed the best results, supporting the potential utility of ensemble learning for ultrasound-based fetal growth assessment, pending further external validation.

Acknowledgements

The authors sincerely thank the Research Department of the Faculty of Medicine for its support.

Footnotes

AI Use Disclosure: For the purpose of Translation, the All Sections Have Been Edited. was used Moderate in the Materials And Methods section.

Authors' Contribution: Study concept/design and supervision: H. S.; Data acquisition: M. M. G. and N. F.; data analysis/interpretation: H. S., M. M. G., and V.D.; Manuscript drafting: M. M. G. and V. D.; Administrative/technical/material support: M. M. G., V. D., N. F., H. S.

Conflict of Interests Statement: The authors do not declare any conflicts of interests for this study.

Data Availability: The dataset presented in the study is available on request from the corresponding author during submission or after publication.

Ethical Approval: This article is derived from a student's master's thesis in Biomedical Engineering at Kermanshah University of Medical Sciences. We hereby express our gratitude and appreciation for the support of the Research Department of the Faculty of Medicine. Furthermore, this article has obtained the ethics code IR.KUMS.MED.REC.1402.108 from the Ethics Committee of Kermanshah University of Medical Sciences.

Funding/Support: No funding was received for this study.

References

- Albu AR, Horhoianu IA, Dumitrascu MC, Horhoianu V. Growth assessment in the diagnosis of fetal growth restriction. Review. *Journal of Medicine and Life*. 2014;7(2):150-4. [PubMed ID: 25408718]. [PubMed Central ID: PMC4197499].
- Wu M, Shao G, Zhang F, Ruan Z, Xu P, Ding H. Estimation of fetal weight by ultrasonic examination. *International Journal of Clinical and Experimental Medicine*. 2015;8(1):540-5. [PubMed ID: 25785028]. [PubMed Central ID: PMC4358483].
- Mgbafulu CC, Ajah IO, Umeora OJ, Ibekwe PC, Ezeonu PO, Orji M. Estimation of fetal weight: A comparison of clinical and sonographic methods. *Journal of Obstetrics and Gynaecology*. 2019;39(5):639-46. [PubMed ID: 31018732]. <https://doi.org/10.1080/01443615.2019.1571567>.
- Seo J, Kim YS. Ultrasound imaging and beyond: Recent advances in medical ultrasound. *Biomedical Engineering Letters*. 2017;7(2):57-8. [PubMed ID: 30603151]. [PubMed Central ID: PMC6208470]. <https://doi.org/10.1007/s13534-017-0030-7>.
- Torrents-Barrena J, Monill N, Piella G, Gratacós E, Eixarch E, Ceresa M, et al. Assessment of radiomics and deep learning for the

- segmentation of fetal and maternal anatomy in magnetic resonance imaging and ultrasound. *Academic Radiology*. 2021;**28**(2):173-88. [PubMed ID: 31879159]. <https://doi.org/10.1016/j.acra.2019.11.006>.
6. Khan W, Zaki N, Masud MM, Ahmad A, Ali L, Ali N, et al. Infant birth weight estimation and low birth weight classification in the United Arab Emirates using machine learning algorithms. *Scientific Reports*. 2022;**12**(1). 12110. [PubMed ID: 35840605]. [PubMed Central ID: PMC9287292]. <https://doi.org/10.1038/s41598-022-14393-6>.
 7. Schmiegelow C, Scheike T, Oesterholt M, Minja D, Pehrson C, Magistrado P, et al. Development of a fetal weight chart using serial trans-abdominal ultrasound in an East African population: A longitudinal observational study. *PLoS ONE*. 2012;**7**(9):e44773. <https://doi.org/10.1371/journal.pone.0044773>.
 8. Kavya C. Performance analysis of different filters for digital image processing. *Turkish Journal of Computer and Mathematics Education*. 2021;**12**(2):2572-6.
 9. Tao J, Yuan Z, Sun L, Yu K, Zhang Z. Fetal birth weight prediction with measured data by a temporal machine learning method. *BMC Medical Informatics and Decision Making*. 2021;**21**(1). 26. [PubMed ID: 33494752]. [PubMed Central ID: PMC7836146]. <https://doi.org/10.1186/s12911-021-01388-y>.
 10. ICC 2019 - 2019 IEEE International Conference on Communications (ICC). 2019.
 11. Kiserud T, Piaggio G, Carroli G, Widmer M, Carvalho J, Neerup Jensen L, et al. The World Health Organization fetal growth charts: A multinational longitudinal study of ultrasound biometric measurements and estimated fetal weight. *PLoS Medicine*. 2017;**14**(1). e1002220. [PubMed ID: 28118360]. [PubMed Central ID: PMC5261648]. <https://doi.org/10.1371/journal.pmed.1002220>.