A Weighted Bag of Visual Word Model for Predicting Fetal Growth Restriction at Early Stage

Yiheng Zhang[1], Xiaowei Huang[1], Mengjie Chen[2], Shiying Huang[2], Weiling Li[1]
1 *School of Computer Science and Technology, Dongguan University of Technology, Dongguan, 523808, China*
2 *Department of Medical Ultrasonics, The Eighth Affiliated Hospital of Sun Yat-sen University, Shenzhen, 518033, China*
15110569705@163.com, huangxw@dgut.edu.cn , chenmj39@126.com, huangshiying0617@163.com, weilinglicq@outlook.com

Abstract: Fetal growth restriction (FGR) is a significant concern for clinicians and pregnant women, as it is a leading cause of morbidity and mortality in newborns. Current research has focused on utilizing ultrasound weight estimation and blood tests to identify FGR, but these methods lack accuracy as they do not consider placental factors. Placental ultrasound images contain potential information for early detection of FGR. but there is currently no effective method to obtain effective knowledge from them. The objective of this study is to achieve the early detection of FGR through the utilization of placental ultrasound images. In this study, we propose a novel visual bag-of-words model for the analysis of placental ultrasound image. Our methodology in-volves the integration of weight scaling within the feature encoding process and the implementation of an ensemble classifier for FGR identification. Through ablation experiment and comparative analysis, we validate the effectiveness of our approach in detecting FGR at early stage.
Keywords: Placenta; Fetal growth restriction; Visual word bag model; Weight scaling; broad learning

I. INTRODUCTION

Fetal Growth Restriction (FGR) refers to the failure to achieve expected physio-logical development which may leads to fetal deformities even fetal death [1] [3]. The prevalence of FGR is on the rise due to the increasing number of pregnancies in advanced maternal age, with global incidence reaching up to 10% in 2021 [4]. Early intervention strategies, such as medication administration during the initial stages of pregnancy, have been shown to mitigate fetal morbidity and reduce risks to maternal health. Therefore, it is particularly important to detect and treat FGR in early pregnancy.

FGR can be caused by multiple factors, e.g., maternal diseases, fetal abnormalities, and placental dysfunction. Notably, research indicates that placental dysfunction is a key factor in the development of FGR [2] [6]. When there is a dysfunction in placental function, notable morphological alterations may manifest. These alterations can be identified using diagnostic imaging modalities like ultra-sound or magnetic resonance imaging, facilitating subsequent investigation and evaluation [19]. **It inspired us if we can learn pattern from placenta ultrasound and associate it with its corresponding fetal FGR results, then we may achieve early detection of FGR.**

FGR detection with placental ultrasound image is essentially an image classification task. In recent years, a variety of image classification models have been developed, with a primary focus on utilizing neural networks. However, neural network models for FGR detection necessitate a substantial volume of training data, a resource that is often constrained in this context. Therefore, a method that can amplify pattern differences on known samples is highly desired.

It is worth noting that, following the direction of healthcare professionals, it has been observed that the distinction between FGR and placental images of healthy fetuses is contingent upon the varying repetition times of regions of interest (ROIs) exhibiting distinct textures. Specifically, placental images of FGR tend to show short rod-like textures, while placental images of normal fetuses typically have dot-like textures. To take advantage of this feature, we came up with the idea of using a visual word bag model.

However, ultrasound imaging exhibits reduced resolution. In early pregnancy, the structural distinctions of placental between FGR and normal fetus are slight, resulting in minimal variations in pattern within ultrasound images. When the texture difference between FGR and normal images is very small, existing BOVW models are difficult to accurately encode them. Moreover, the long gestational period results in a limited number of data samples. Hence, a method which is able to maximize learning patterns from a small number of samples and accurately associate them with corresponding FGR results is desired.

To address above issues, this study proposes a weighted bag of visual word (WBOVW) model for detecting fetal growth restriction at early stage with two-fold ideas:

a) encoding placental ultrasound with a novel weighted bag of visual word model;

b) learning association knowledge of the encoded placental ultrasound image and its corresponding FGR result using ensemble classifiers.

The contribution of this work is as follows:

a) achieving FGR detection at early pregnancy with only placental ultrasound images;

b) providing a weighted bag of visual word method.

Experimental results on real ethically certified data collected by a hospital indicate that the proposed method has the ability to detect FGR at early stage.

II. RELATED WORK

FGR Detection: Various indirect approaches have been suggested to detect FGR. For instance, a study outlined in reference [5] detects FGR by analyzing the umbilical artery blood flow indices with mathematical models. Yan Feng et al. utilized early and mid-pregnancy biophysical features for detecting FGR [7]. How-ever, these approaches are either incapable of identifying FGR during its initial phases or depend on the biological characteristics obtained through costly equipment.

Image Classification: Numerous image classification models have been introduced in recent years, predominantly leveraging neural networks. For instance, He et al used Res-Net won the ImageNet competition championship in 2015 [8]. Tan et al. proposed a lightweight model, Efficient-Net performed well on the CIFAR-100 dataset with fewer parameters [9]. The ViT (vision transformer) pro-posed by Google directly applies the transformer to the model of image classification [10]. However, neural network classification models are not suitable for this type of medical small sample task.

Bag of Visual Words (BOVW): Compared to deep learning, BOVW's manual feature extraction method is more suitable for small sample classification. Many researchers have made improvements based on BOVW. Sun et al. add location information to the model and use speed up robust features (SURF) and color vector angle histogram (CVAH) for local feature description, achieving results in product classification testing [11]. Zahra et al. proposed a model composed of bags of visual words and neural net-work classifiers, which achieved results in the detection task of COVID-19 [12]. However, existing BOVW models cannot accurately encode placental ultrasound images. In this article, we propose a weight scaling method that can effectively encode.

III. METHOD

The proposed WBOVW model, illustrated in Figure 1, comprises two primary com-ponents: the establishment of a visual word bag library and image encoding classification. Construction of visual word bag library: We will use training set images to build this visual word bag library. Note that the input of WBOVW is placental ultrasound images. In each image, a ROI is marked by imaging physicians. The ROIs are segmented from images as the preprocessed image and feed to the key points feature extraction module. It selects key-points from ROI with Scale Invariant Feature Transform (SIFT) and learns features of each key-point with Histo-grams of Oriented Gradients (HOG). Afterwards, Gaussian Gixture Model (GMM) is used to construct visual words bag based on the extracted image block features of each ROI. Image encoding classification: This part will use a visual word bag library to feature encode all images. Each image undergoes the same preprocessing and feature extraction stages as mentioned above. The extracted features of each image will be calculated for similarity with visual words and encoded with weight scaling. The encoded image features can be trained through a classifier, and finally, ensemble prediction is used to improve classification accuracy.

A. Key Feature Representation

Keypoints selection with SIFT keypoint detection. Scale-invariant feature transform (SIFT) is an efficient method for finding key points on different scale spaces [13]. It is proven to be robust. Note that ROI, which is marked by data provider, in placental ultrasound image are with different size, the accuracy of features extracted directly from image is affected by the size of image and the semantic order of features. Therefore, we first employ SIFT keypoint detection to identify keypoints in each image.

The SIFT algorithm first applies a Gaussian filter to the image for filtering:

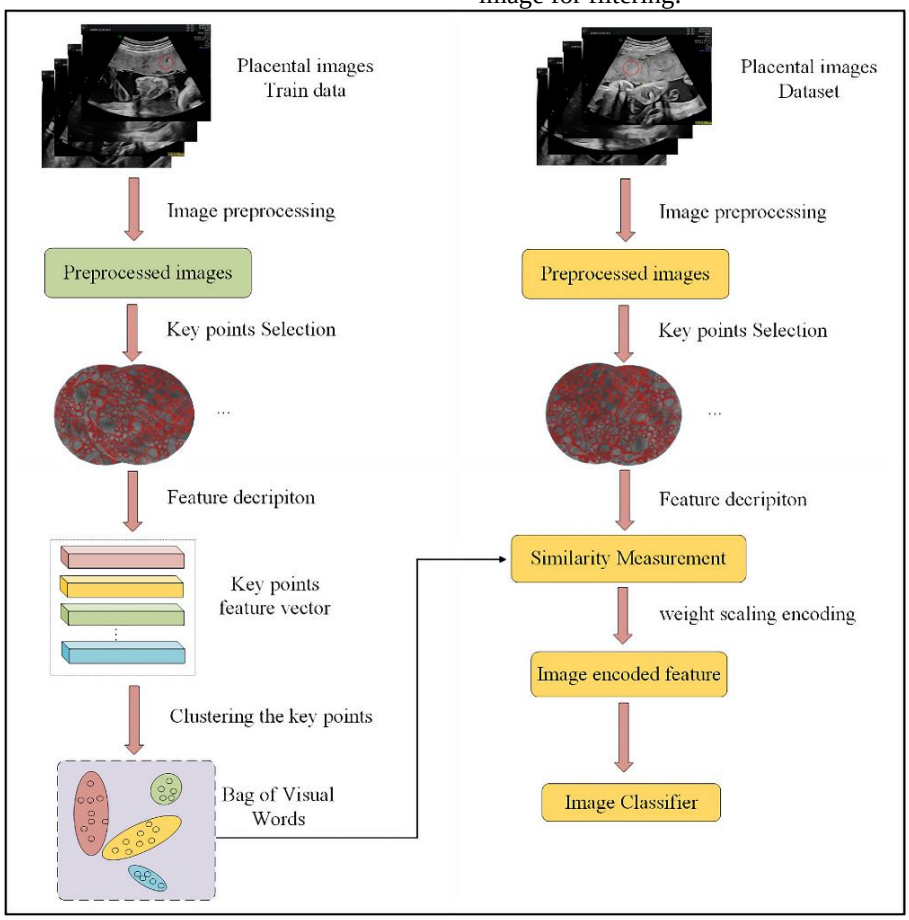

Figure 1. The overall framework of our method

$$L(x,y,\sigma) = G(x,y,\sigma) \otimes I(x,y), \tag{1}$$

where $I(x,y)$ is the input image, $G(x,y,\sigma)$ is the Gaussian filter, $\otimes$ represents filtering operations and $L(x, y, \sigma)$ is the filtered image. The Gaussian filter is defined as follows:

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-m/2)^2+(y-n/2)^2}{2\sigma^2}}, \tag{2}$$

where σ represents the standard deviation of the Gaussian distribution, m and n denote the size of the Gaussian filter, x and y represent the positions of the corresponding elements.

In order to find the extreme points of the image, by subtracting the adjacent image matrices in the same scale space, the Gaussian difference scale space can be obtained as follows:

$$D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y,\sigma), \tag{3}$$

where k is the scale factor. The Difference of Gaussians (DoG) model is illustrated in Figure 2.

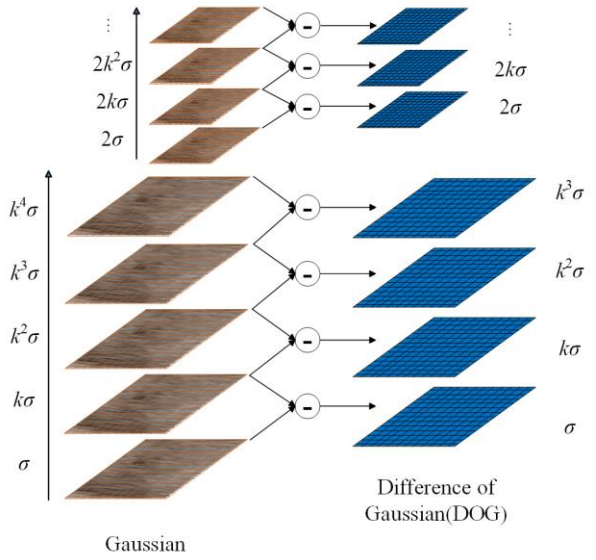

Figure 2. An overview of DOG

After establishing the DoG, we search for local extrema points. Each pixel is compared with its neighbors to check if it is larger or smaller than all its neighbors. The adjacent point is composed of 8 adjacent points of the same scale and 9×2 points of adjacent scales up and down, a total of 26 points. This results in key points at different scales in the image, increasing diversity in BOVW construction.

Keypoints Feature extraction with HOG. Select domains of different sizes around each key point based on their scale size, and perform HOG feature extraction on this domain [14]. Firstly, divide the domain into 4x4=16 sub regions and calculate the pixel gradient size and direction in each sub-region.

$$G_x(x,y) = I(x+1,y) - I(x-1,y), \tag{4}$$

$$G_y(x,y) = I(x,y+1) - I(x,y-1), \tag{5}$$

$$G(x,y) = \sqrt{G_x^2 + G_y^2}, \tag{6}$$

$$\varphi(x,y) = \arctan\frac{G_y(x,y)}{G_x(x,y)}, \tag{7}$$

where $G(x,y)$ denotes the gradient size representing pixels, $\Phi(x,y)$ the gradient direction of pixels.

Then, we divide the range of 0°-180° into 9 intervals: [0°, 20°], [20°, 40°], ..., [160°, 180°]. For each sub region, the gradient size is accumulated and normalized based on the interval corresponding to the pixel gradient direction value.

Finally, the feature descriptors of 16 sub regions were concatenated together to obtain $V_{i,j}$, representing the feature vector of key point j in image i.

B. Encoder

Building the bag of visual word with GMM. Through key feature representation component, we obtained feature matrix $V_i \in R^{mi \times n}$ for i-th placenta ultrasound image, where m_i represents the number of key points in image i, and n represents the feature vector of the key points. With V_i, the key feature of the i-th placenta ultrasound image can be represented. However, it cannot be utilized for prediction directly since a) V_i has different size; b) the order of ROIs recorded by V brings additional spatial semantic information.

In order to fully utilize the key features, a method which can recode them is required. BOVW has proven to be effective in recoding features, traditional Bag-of-Visual-Words (BOVW) methods typically use k-means clustering to cluster the feature vectors. However, in the paper, the clustering model employed is the GMM Model [15], which not only incorporates the distance information of the sample points but also includes information about the number of sample points and their variances. We use GMM to cluster the keypoints of all images in the training set to construct BOVW.

In GMM, it is assumed that the data is composed of several Gaussian distributions. The formula for the Gaussian mixture model is:

$$\mathrm{P}(V_{i,j}) = \sum_{k=1}^{K} \pi_k N(V_{i,j} \mid u_k, \Sigma_k), \tag{8}$$

where u_k denotes the mean factor of Gaussian distribution, $\sum_k$ is the covariance, π_k is the weight of each Gaussian model, K is the number of Gaussian models.

Note that the goal of GMM is to maximize the likelihood function as follows:

$$L = \prod_{i=1}^{I} \prod_{j=1}^{Ji} P(V_{i,j}), \tag{9}$$

where I is the number of images, J_i the number of key points in i-th image. Hence, these parameters in (8), i.e., u_i, $\sum_i$ and π_i, can be derived with Expectation Maximization (EM) algorithm. By implementing forementioned process, a dictionary can be obtained for supporting image encoding.

Encoding input placental ultrasound image. Recall that we use the GMM model to calculate the probability of key points belonging to each cluster, which includes distance information, covariance information, and weight information. Ideally, each key point should be assigned to a specific cluster, which means that the probability of belonging to a specific cluster class should be relatively high. However, due to the low resolution of medical images and the influence of noise, some key points may fall at the intersection of certain clusters. It belongs to different cluster classes with similar probabilities, and weighted statistics based on this method may not be conducive to classification. To reduce the impact of noise, we propose a weight scaling statistical method in this work. The feature encoding part is shown in Figure 3.

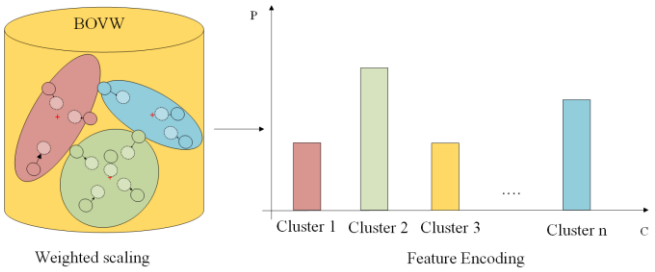

Figure 3. Feature encoding

Firstly, we use the trained GMM model to perform probability prediction on each key point in the image.

$$p_{i,j,k} = \pi_k N(V_{i,j} \mid u_k, \Sigma_k), \tag{10}$$

where $p_{i,j,k}$ is the probability of $V_{i,j}$ belongs to the k-th cluster.

Identify the labels of each cluster to which the key points belong based on probability values. Let $P_{i,j}$ be a length K vector, e.g., $[p_{i,j,1},\ldots, p_{i,j,k},\ldots, p_{i,j,K}]$. A length K vector, e.g., $Q_{i,j}=[q_{i,j,1},\ldots, q_{i,j,k},\ldots, q_{i,j,K}]$, is proposed to encode the j-th key point of i-th image, where $q_{i,j,k}\in\{0, 1\}$ and $\sum q_{i,j,k}=1$. Moreover, $q_{i,j,k}=1$ if $p_{i,j,k}$ is the maximum among all ps. Then, the weight scaling formula is as follows:

$$U_i = \frac{\sum_{j=1}^{J} \alpha * P_{i,j} + (1-\alpha) * Q_{i,j}}{J}, \tag{11}$$

where U_i represents the re-encoded features of image i, and α is a weight scaling hyperparameter.

The above approach can cause feature points to move towards the corresponding cluster center. There are significant differences in histogram statistics for key points located at certain cluster boundaries. When α=1, it becomes a probability-weighted average, and when α=0, it becomes a class-weighted statistic. Introducing the α hyperparameter allows for better adjustment, depending on different datasets and their specific requirements. After feature encoding, each image has a corresponding feature vector U_i with same dimension.

C. FGR Predictor

Currently, many models use fully connected neural networks for classification. However, when the amount of data is small, there is a risk of overfitting. Therefore, it is not suitable for tasks like FGR prediction. In machine learning, Support Vector Machines (SVM) use "kernel functions" to map data from the original feature space to a higher-dimensional feature space, making the data linearly separable in that space. It can be used to solve binary classification tasks Besides, BLS involves mapping input features through nonlinear transformations to a higher-dimensional feature space where data becomes more easily separable and classifiable [16][18]. It can be seen as a variant of fully connected neural networks, but while fully connected neural networks focus on constructing deep architectures, BLS focuses on constructing wide architectures, as shown in Figure 4.

BLS does not directly input the original feature vectors into the network. Instead, it first expands the original features through a feature mapping layer.

$$Z_{zi} = \phi(U_i * W_{zi} + \beta_{zi}), zi = 1....n, \tag{12}$$

where W_{zi} and β_{zi} are parameters, Φ represents the non-linear transformation, and Z_{zi} represents the l-th node after the mapping node.

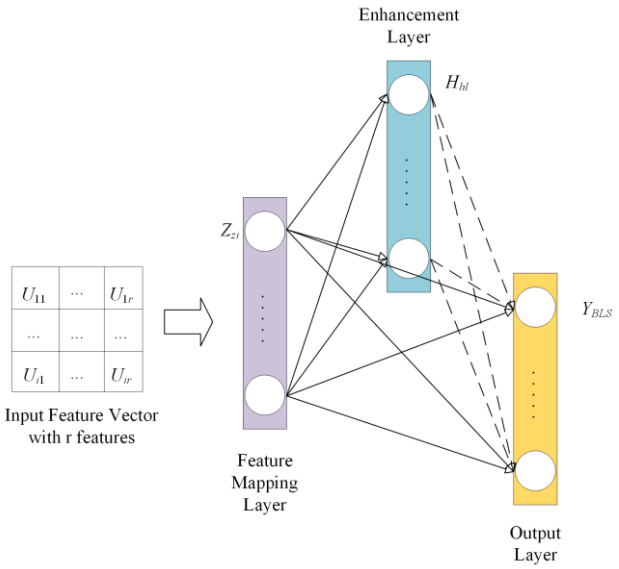

Figure 4. Structure of Broad Learning System

The nodes obtained in the previous step are then passed through an enhancement layer to obtain H_l. Finally, Z_l and H_l are concatenated, and the concatenated result is passed through linear and non-linear transformations to output the predicted values.

$$H_{hl} = \xi([Z_1 Z_2 ... Z_n] * W_{hl} + \beta_{hl}), hl = 1,....m, \tag{13}$$

where ζ represents the non-linear transformation, W_{hl} and β_{hl} are parameters.

$$Y_{BLS} = [Z_1,...Z_n \mid H_1,...H_m]W^m = AW^m, \tag{14}$$

where W^m represents the parameters, and Y_{BLS} represents the predicted probabilities from the BLS classifier.

The parameter W^m can be obtained through optimization and inverse matrix solution.

$$\arg\min_{W^m} : \| AW^m - Y_{BLS} \|_2^2 + \varepsilon \| W \|_2^2, \tag{15}$$

$$W^m = (\varepsilon I + AA^T)^{-1} A^T Y_{BLS}, \tag{16}$$

where ε is regularization parameter.

In broad learning, as the width increases, the risk of overfitting becomes more severe, and a larger amount of data is required. In fact, the accuracy of an ensemble of multiple smaller models may be higher than that of a single large model. BLS and SVM adopt different mapping methods to enhance data separability. In this paper, we integrate these two classifiers for predictive modeling in medical tasks, aiming to improve the robustness of the model. The SVM classifier can record the distance of each sample to the hyperplane. We use the sigmoid function to convert this distance into a probability value. On the other hand, the last layer of the BLS outputs the score values for each class for each sample. We normalize the output values using softmax to convert them into probability values. Finally, we can use λ and 1-λ as weights to combine the probability values of the two models. The resulting combined probability is used for ensemble prediction. By adjusting the hyperparameter λ, we can control the mixing ratio of the two models.

$$Y_{i,j} = \frac{\exp(Y_{i,j}^{BLS}) * \lambda}{\sum_{j=0}^{1} \exp(Y_{i,j}^{BLS})} + \frac{1*(1-\lambda)}{1+\exp(-dist(Y_{i,j}^{SVM}, L))}, \tag{17}$$

Table 1. Results of ablation experiment

No.	Feature extraction			Feature coding			Classifer			Accuracy
	LBP	Harr-like	Hog	Kmeans	GMM	GMM-scaled	SVM	BLS	BLS+SVM	
E1	√					√			√	60.67%
E2		√				√			√	65.57%
E3			√	√					√	64.07%
E4			√		√				√	68.52%
E5			√			√	√			66.21%
E6			√			√		√		60.34%
E7			√			√			√	**70.68%**

In the formula, $Y_{i,j}^{BLS}$ represents the score assigned by the BLS classifier to sample *i* for class *j*, -dist($Y_{i,j}^{SVM}$,L) represents the distance from the hyperplane when sample *i* is predicted as class *j* by SVM, L represents the constructed hyperplane by SVM, and $Y_{i,j}$ represents probability of sample *i* belonging to class *j*.

IV. EXPERIMENT

A. Experiment settings

Dataset. An ethically reviewed placental ultrasound image dataset is utilized in this work. It contains 80 cases from a hospital where fetuses were at maturity level 1 and were between 18 to 24 weeks of gestation. Expert clinicians tagged the ROIs in these images, as shown in Figure 5. Among them, 40 are FGR images and 40 are normal fetal images. In order to evaluate the performance of involved models, we used cross validation method, with a ratio of 6:2:2 for training set to validation set to test set. This experiment has been repeated at least 10 times or more. All outcome indicators are obtained from the test set.

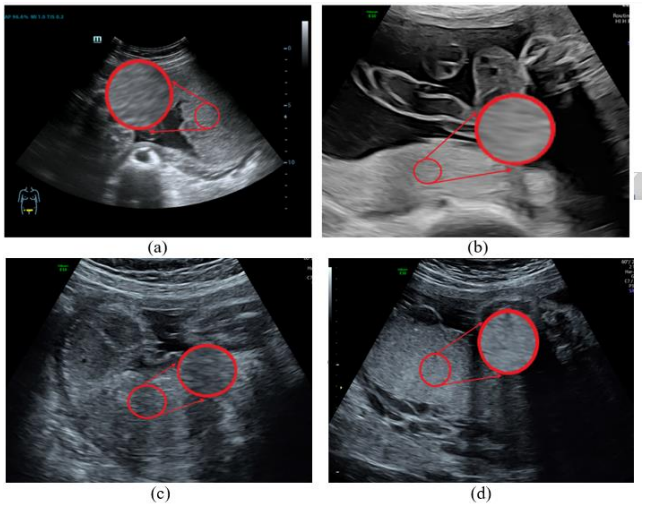

Figure 5. Sample of placental ultrasound images. (a) and (c) are placental ultrasound images with FGR, (b) and (d) are normal placental ultrasound images.

Implement details. In image preprocessing, wavelet transform is utilized for denoising. Due to the unstable position and angle of the ultrasound probe during image acquisition, as well as the small image dataset, we also adopt brightness variation, contrast trans-formation, random rotation and random cropping as augmentation methods for each image. Note that we did not use operations such as random scaling and affine transformation, as this would disrupt the original texture details of the placental image. To enhance the texture of the images, we used histogram equalization for texture enhancement. In terms of parameter settings, we set the number of clusters in the visual word bag model is 100(k=100). In the weight scaling stage, we set α=0.15, and in the joint detection stage, we set λ=0.3.

Evaluation Metrics. To evaluate the performance classification, we adopt Accuracy, Recall and Precision.

$$Accuracy = \frac{TN + TP}{FP + FN + TP + TN}, \tag{18}$$

$$Precision = \frac{TP}{FP + TP}, \tag{19}$$

$$Recall = \frac{TP}{TP + FN}, \tag{20}$$

where *TP* denotes the number of true positive cases in results, *FP* the number of false positive cases. *FN* the number of false negative cases. and *TN* the number of true negative cases.

B. Ablation Experiment

In order to understand the effectiveness of each component of the proposed model, we first conduct an ablation experiment. The experimental results are listed in Table 1. From it, we can conclude that each component has positive contribution to the final prediction accuracy. For instance, the Accuracy of proposed model, i.e., E7, is 0.707, which is 10.4% higher than accuracy at 0.603 by E6, 4.5% higher than accuracy at 0.662 by E5. This indicates that the integrated prediction of BLS and SVM is superior to a single prediction. E7's Accuracy is 2.2% higher than accuracy at 0.685 by E4, 6.7% higher than accuracy at 0.640 by E3. This indicates that our proposed weighted encoding method is effective in processing ultrasound images. This encoding method reduces the impact of ultrasound noise. E7's Accuracy is 5.1% higher than accuracy at 0.656 by E2, 10.1% higher than accuracy at 0.606 by E1. Ablation experiment results indicate that our proposed method is effective.

C. Comparison against state-of-the-art models

As shown in Table 3, the proposed model, i.e., M7, obtains higher Accuracy, Recall and Precision than its peers. For instance, M7's Accuracy is 0.707, which is 4.2% higher than accuracy at 0.665 by M6, 6.7% higher than accuracy at 0.640 by M5, 8.7% higher than accuracy at 0.620 by M4, 12.6% higher than accuracy at 0.581 by M3, and 14.9% higher than accuracy at 0.558 by M2, 17.2% higher than accuracy at 0.535 by M1. In addition, M7 has shown better excellence in

Precision. Experiment results indicate that proposed models have evident advantage in various indicators, which denotes that the proposed Weighted Bag of Visual Word Model has high accuracy of predicting FGR.

Table 2. Involved models

No.	Description
M1	The vision transformer model for image recognition at scale pretrained by imagenet dataset [10]
M2	The deep residual network pretrained by imagenet dataset.[8]
M3	The EfficientNet pretrained by imagenet dataset [9]
M4	The improved bag of visual words model for commodity image classification [11]
M5	The improving bag of visual words model via combining deep features with feature difference vector [17]
M6	The bag of visual words using neural networks for detection of COVID-19 in X-ray [12]
M7	The FGR prediction model proposed in this work.

Table 3. Comparison of experimental results

Model	*Accuracy*	*Recall*	*Precision*
M1	53.48%	47.45%	50.46%
M2	55.75%	50.63%	49.34%
M3	58.17%	53.00%	52.65%
M4	62.06%	62.50%	38.46%
M5	64.07%	62.25%	57.64%
M6	66.51%	63.53%	58.53%
M7	**70.68%**	**66.28%**	**71.23%**

D. Summary

Predicting FGR through placental ultrasound images is essentially an image classification task. However, the noise caused by ultrasound leads to inaccurate representation. The experiment shows that our proposed encoding mode can reduce the impact of noise and improves classification accuracy. This model for predicting FGR through images has the feasibility of clinical application. Assisting doctors in auxiliary diagnosis can reduce their workload. In some areas where medical equipment is underdeveloped, this application has higher value. Obtaining cross-sectional images from ultrasound video streams and delineating ROI in the images requires professional expertise. We can find a method in subsequent work to assist doctors in delineating ROI and searching for cross-sectional areas. In addition, we can also combine placental image information and other biometric information for multimodal prediction of FGR.

V. CONLUSION

In order to predict Fetal growth restriction at early stage, we propose a visual bag-of-words model based on weight scaling for analyzing placental ultrasound images of pregnant women in the early stages of pregnancy. Specifically, we introduce weight scaling during the feature encoding stage and design an ensemble classifier for FGR prediction. Through ablation experiments and comparative analysis, we validate the effectiveness of our approach, achieving an accuracy of 70% in predicting FGR using limited samples. The proposed model show its potential to assist doctors in making preliminary judgments at early pregnancy, which greatly helps with treatment and allows patients to receive treatment as soon as possible, reducing the harm to pregnant women.

ACKNOWLEDGMENT

This research is supported in part by the National Natural Science Foundation of China under grants 62102086, in part by the Guangdong Basic and Applied Basic Research Foundation under grants 2022A1515140102 and 2021B1515140046. (Corresponding Authors: Weiling Li and Xiaowei Huang)

REFERENCES

[1]. Shrivastava, D., Master.: A. Fetal Growth Restriction. J Obstet Gynecol India 70, 103–110 (2020). https://doi.org/10.1007/s13224-019-01278-4

[2]. M.L.E. Hendrix, J.A.P. Bons, N.O. Alers, et al.: Maternal vascular malformation in the placenta is an indicator for fetal growth restriction irrespective of neonatal birthweight. Placenta,Volume 87,2019, Pages 8-15, ISSN 0143-4004, https://doi.org/10.1016/j.placenta.2019.09.003.

[3]. Debora F. B. Leite, Jose G. Cecatti.: Fetal Growth Restriction Prediction: How to Move beyond. The Scientific World Journal, vol. 2019, Article ID 1519048, 8 pages, 2019. https://doi.org/10.1155/2019/1519048

[4]. Fetal Growth Restriction: ACOG Practice Bulletin, Number 227. Obstetrics & Gynecology 137(2):p e16-e28, February 2021. https://doi.org/10.1097/aog.0000000000004251

[5]. Xinying Yu, Ye Yao, Dan Wang, Jiani Tang, Jing Lu.: Prediction of Fetal Growth Restriction for Fetal Umbilical Arterial/Venous Blood Flow Index Evaluated by Ultrasonic Doppler under Intelligent Algorithm. Computational and Mathematical Methods in Medicine, vol. 2022, Article ID 7451185, 8 pages, 2022. https://doi.org/10.1155/2022/7451185

[6]. Salomon LJ, Alfirevic Z, et al.: ISUOG Practice Guidelines: ultrasound assessment of fetal biometry and growth. Ultrasound Obstet Gynecol. 2019 Jun;53(6):715-723. https://doi.org/10.1002/uog.20272

[7]. Feng Y, Zheng H, Fang D, Mei S, Zhong W, Zhang G.: Prediction of late-onset fetal growth restriction using a combined first- and second-trimester screening model. J Gynecol Obstet Hum Reprod. 2022 Feb;51(2):102273. https://doi.org/10.1016/j.jogoh.2021.102273

[8]. K. He, X. Zhang, S. Ren and J. Sun.: Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778. https://doi.org/10.1109/CVPR.2016.90

[9]. Tan M, Le Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. International conference on machine learning. PMLR, 2019: 6105-6114.

[10]. A. Dosovitskiy and Lucas Beyer et al.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. International Conference on Learning Representations (2020):abs/2010.11929

[11]. Huadong Sun, Xu Zhang, Xiaowei Han, Xuesong Jin, Zhijie Zhao.: Commodity Image Classification Based on Improved Bag-of-Visual-Words Model. Complexity, vol. 2021, Article ID 5556899, 10 pages, 2021. https://doi.org/10.1155/2021/5556899.

[12]. Zahra Nabizadeh-Shahre-Babak, Nader Karimi, Pejman Khadivi, Roshanak Roshandel.: Detection of COVID-19 in X-ray images by classification of bag of visual words using neural networks. Biomedical Signal Processing and Control,Volume 68,2021,102750,1746-8094, 1, https://doi.org/10.1016/j.bspc.2021.102750.

[13]. D. G. Lowe.: Object recognition from local scale-invariant features. Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 1999, pp. 1150-1157 vol.2, doi: https://doi.org/10.1109/ICCV.1999.790410.

[14]. N. Dalal and B. Triggs.: Histograms of oriented gradients for human detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 886-893 vol. 1, https://doi.org/10.1109/CVPR.2005.177

[15]. P. Johnston, K. Nogueira and K. Swingler.: GMM-IL: Image Classification Using Incrementally Learnt, Independent Probabilistic Models for Small Sample Sizes. in *IEEE Access*, vol. 11, pp. 25492-25501, 2023, doi: https://doi.org/10.1109/ACCESS.2023.3255795

[16]. L. P. Chen and Z. Liu.: Broad Learning System: An Effective and Efficient Incremental Learning System Without the Need for Deep Architecture. IEEE Transactions on Neural Networks and Learning

Systems, vol. 29, no. 1, pp. 10-24, Jan. 2018. https://doi.org/10.1109/TNNLS.2017.2716952

[17]. X. Wang.: Improving Bag-of-Deep-Visual-Words Model via Combining Deep Features With Feature Difference Vectors. IEEE Access, vol. 10, pp. 35824-35834, 2022, https://doi.org/10.1109/ACCESS.2022.3163256

[18]. K. Yang, Z. Yu, C. L. P. Chen, W. Cao, J. You and H. -S. Wong.: Incremental Weighted Ensemble Broad Learning System for Imbalanced Data," in IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 12, pp. 5809-5824, 1 Dec. 2022, https://doi.org/10.1109/TKDE.2021.3061428

[19]. Burton G J, Jauniaux E.: Pathophysiology of placental-derived fetal growth restriction[J]. In: American journal of obstetrics and gynecology, 218(2): S745-S761(2018). https://doi. org/10. 1016/j. ajog. 2017. 11. 577