Published in final edited form as:

J Magn Reson Imaging. 2018 February ; 47(2): 449–458. doi:10.1002/jmri.25806.

In vivo placental MRI shape and textural features predict fetal growth restriction and postnatal outcome

Sonia Dahdouh, PhD¹, Nickie Andescavage, MD^{1,2,5}, Sayali Yewale, MD¹, Alexa Yarish, BS¹, Diane Lanham, MS¹, Dorothy Bulas, MD⁴, Adre J du Plessis, MD^{3,5}, and Catherine Limperopoulos, PhD^{1,4,5}

¹Developing Brain Research Laboratory, Children's National Health System, Washington DC, USA

²Division of Neonatology, Children's National Health System, Washington DC, USA

³Fetal & Transitional Medicine, Children's National Health System, Washington DC, USA

⁴Diagnostic Imaging & Radiology, Children's National Health System, Washington DC, USA

⁵Department of Pediatrics, George Washington University School of Medicine, Washington DC, USA

Abstract

Purpose—To investigate the ability of 3D MRI placental shape and textural features to predict fetal growth restriction (FGR) and birth weight (BW) for both healthy and FGR fetuses.

Materials and Methods—We recruited two groups of pregnant volunteers between 18 and 39 weeks of gestation; 46 healthy subjects and 34 FGR. Both groups underwent fetal MR imaging on a 1.5T GE scanner using an 8-channel receiver coil. We acquired T2-weighted images on either the coronal or the axial plane to obtain MR volumes with a slice thickness of either 4 or 8mm covering the full placenta. Placental shape features (volume, thickness, elongation) were combined with textural features; first order textural features (mean, variance, kurtosis and skewness of placental grey levels), as well as, textural features computed on the grey level co-occurrence and run-length matrices characterizing placental homogeneity, symmetry and coarseness. The features were used into two machine learning frameworks to predict FGR and BW.

Results—The proposed machine-learning based method using shape and textural features identified FGR pregnancies with 86% accuracy, 77% precision and 86% recall. BW estimations were $0.3 \pm 13.4\%$ (mean percentage error $\pm$ standard error) for healthy fetuses and $-2.6 \pm 15.9\%$ for FGR.

Conclusion—The proposed FGR identification and BW estimation methods using *in utero* placental shape and textural features computed on 3D MR images demonstrated high accuracy in our healthy and high-risk cohorts. Future studies to assess the evolution of each feature with regard to placental development are currently underway.

Corresponding Author: Catherine Limperopoulos, Ph.D., Children's National Health System, 111 Michigan Ave. NW, Washington, D.C. 20010, climpero@childrensnational.org, Tel: 202-476-5293, Fax: 202-476-6833.

The authors declare that they do not have any competing financial interests.

Keywords

Placenta; Fetal Growth Restriction; MRI; Textural analysis; Shape analysis

INTRODUCTION

The human placenta is a critical organ for the developing fetus, and has vital roles in the provision of necessary nutrients and the excretion of toxic by-products. In addition to supporting the metabolic demands of the growing fetus, the placenta also provides important immune-modulating and endocrine roles (1). When placental support begins to fail, the resulting placental dysfunction can adversely impact both maternal and fetal health, as seen in pre-eclampsia, fetal growth restriction (FGR), preterm birth, and fetal demise (2). Moreover, surviving infants are at risk for lifelong cardiovascular, metabolic and neuropsychiatric sequelae (3).

Despite the important role of placental development in maternal-fetal health, there is a paucity of non-invasive tools to assess placental function *in utero*. Indeed, placental disease may not be suspected until well after complications of fetal growth have occurred, and in some cases may not be identified until after delivery (3). Noteworthy, up to 45% of infants born less than the tenth percentile for weight (i.e., small for gestational age (SGA)), are not detected until after delivery (4). Furthermore, placental disease that results in impaired growth trajectories in fetuses and newborns with weights greater than the tenth percentile (i.e., appropriate for gestational age (AGA)) represent another high-risk group that currently goes undetected (5).

Fundal height and ultrasonography are the primary screening tools to assess fetal growth, yet remain insensitive for the detection and diagnosis of fetal growth failure, particularly for infants at the lower end of the growth spectrum (6). Sonographic estimations of fetal weight at the time of delivery and prediction of birth weight (BW) can vary by over 15% (6). Moreover, for SGA fetuses, estimation of fetal weight is even more problematic, with sensitivity as low and variable as 35 – 57% (7). Alternate methods for predicting BW include *in vivo* two-dimensional (2D) and three-dimensional (3D) sonography of placental volume and surface area (8, 9). Additional qualitative descriptions of both *in vivo* and *ex vivo* placental morphometry from routine sonography are also associated with BW and FGR (10), however these methods have not been incorporated into routine clinical care.

Quantitative descriptions of placental development using routine sonography and textural analysis have also shown to correlate with gestational age, but have not been studied relative to fetal growth (11). Very few studies have applied alternate imaging modalities, such as magnetic resonance imaging (MRI) to quantify placental development in healthy as well as growth restricted pregnancies. These limited placental MRI studies were based on 2-D measurements and have described abnormalities in placental volume, thickness and intensity in high-risk pregnancies (12, 13). Damodaram *et al.* (13) highlighted significant correlations between FGR and placental volume as well as placental thickness to volume ratio in FGR. They also observed different trajectories for placental appearance between FGR and healthy fetuses. However, complex 3-D modeling of *in vivo* placental development, including

volume, morphometry and textural analysis is largely unexplored. Moreover, the sensitivity of these 3-D MRI measures for reliably diagnosing fetal-neonatal growth has not been studied.

Capitalizing on the different findings presented above, we investigated the ability of in vivo structural MRI placental features to (1) identify FGR and (2) estimate BW in both healthy and high-risk pregnancies using a novel semi-automated framework derived solely from placental measures.

MATERIAL AND METHODS

Subjects

Pregnant women were recruited prospectively as part of an ongoing longitudinal, observational study which was approved by the hospital's Institutional Review Board and written informed consent was obtained from all participants.

We studied a total of 80 pregnant women in which we acquired a total of 124 fetal scans. The research protocol initially included a single fetal MRI acquisition at any time point after 18 weeks of gestation and was later amended to include a second fetal MRI scan when feasible.

We recruited two groups of pregnant women. The control group consisted of 46 pregnant volunteers with healthy fetuses (36 of whom had two fetal MRI scans). Only singleton pregnancies with normal maternal health histories, normal prenatal testing and no contra-indication for fetal MRI were included. Health status was determined at the time of recruitment and subsequently re-confirmed at birth based on the uncomplicated, term delivery of an AGA newborn.

The second group, recruited as part of an ongoing prospective study on placental disease and fetal brain development, consisted of 34 women with pregnancies complicated by FGR (8 of which had two fetal MRI scans). FGR was defined when the estimated fetal weight was below the tenth percentile for GA (14) secondary to suspected placental insufficiency (abnormal Doppler of the fetal vessels or evidence of asymmetric growth with lagging abdominal circumference). The second fetal MRI was deferred if FGR was not identified until > 32 weeks and/or not feasible if the mother delivered prematurely. Multiple gestation pregnancy, pregnancies with known or suspected congenital infections as well as known or suspected genetic abnormalities were excluded in both groups.

Table 1 summarizes the baseline characteristics of our cohort.

Data Acquisitions

Single shot fast spin echo (SSFSE) T2-weighted images were acquired on a 1.5T Discovery MR450 scanner (GE Healthcare, Milwaukee, Wisconsin) using an 8-channel surface receiver coil. Two sets of acquisition parameters were used: TE=160ms, TR=1100ms, 4mm slice thickness and 40 to 60 consecutive slices for full placental coverage in either the maternal coronal or axial plane for a final in-plane resolution of 1.64 mm for the first set,

and TE=160ms, TR=1100ms, 8mm slice thickness in the maternal coronal plane for a final in-plane resolution of 0.85. No contrast or sedation was used for any of the MRI studies.

Data Pre-processing

All placental volumes were manually segmented in the plane of acquisition and then further corrected on the other planes to ensure spatial consistency. The segmentations were performed by an annotator (AY, SY) with more than one year of experience in MRI placental segmentation and reviewed by a senior neonatologist (NA) with more than five years of experience. All were blinded to FGR vs. control status.

Shape Descriptors

Using the placental segmentations, 3D meshes were reconstructed using a meshing framework developed in-house based on the Computational Geometry Algorithms Library (CGAL) (15) which proposes a set of algorithms to reconstruct, manipulate and analyze surfaces and volumetric meshes. A Poisson surface reconstruction (16) was then performed on the nodes of the initial mesh to obtain the final 3D surface. To characterize placental shape, three different 3D shape features were used: volume, thickness and elongation. Indeed, geometric features, such as volume and thickness, are the clinical standard placental morphometric analyses for both US and MRI (17). Moreover, alteration of volume and thickness to length ratio has been previously described in FGR fetuses (17).

1. Volume was defined as the 3D volume of the placental mesh.
2. Elongation was defined as the length of the longest branch of the 3D medial axis skeleton of the shape. It was computed on the binary volume of the placenta based on the homotopic thinning algorithm (18) and its implementation in (190).
3. Thickness was defined as the maximal distance existing between the points of the placenta belonging to the maternal side and their projection on the fetal side. Proposed as a mesh descriptor for the segmentation of 3D articulated shapes, the *Shape Diameter Function* computes the diameter of the placental volume at each node of the surface mesh based its neighborhood (20). The non-normalized maximal value of this scalar function is used as a measurement of the placenta thickness. Computation is based on the CGAL implementation of the descriptor.

Figure 1 illustrates the different features presented above.

Textural Features

Three sets of textural features computed on the whole placenta, were used to characterize placental grey level appearance.

The first set, based on first order statistics, was used as a macroscopic characterization of placental texture. It was composed of the mean, variance, kurtosis and skewness of the grey levels distribution of the placenta. To account for intensity variation between different images, these metrics were normalized by their respective values on the whole image.

The second set, computed on the well-known Grey Level Co-occurrence Matrix (GLCM) (21, 22) was composed of the energy, entropy, inverse difference moment, contrast, cluster shade and cluster prominence. These features allow for a more refined characterization of the placental texture as they capture spatial dependencies between pairs of grey levels. Each feature was averaged over all directions. The GLCM matrix was computed on a reduced number of grey levels: 128 for FGR identification and 32 for BW estimation.

To further characterize placental texture, a third set of ten features based on the run-length statistics (23 – 25) was computed. While most run-length features are often difficult to clinically interpret as a whole, they can be used to describe the coarseness of a texture. More specifically, the run-length matrix computes the number of “runs” or vectors of consecutive voxels having the same gray level in a given orientation (24). A texture presenting a high number of short runs could then be interpreted as finer with more details than a texture with a high number of long runs (23, 26). These run-length based features were computed on the same grey level reduction as above.

Table 2 summarizes the main characteristics of the textural features based on the GLCM and Run-length matrices used.

Prediction Framework

In Utero Detection of FGR

Placental signature: For each placenta, shape and textural features were computed and a global placental signature was created that included only these features in addition to GA at the time of the MRI study and gender. Since no single feature was likely to be able to accurately predict outcome on its own, we used the combination of all the descriptors to better delineate placental development, identify FGR and predict BW.

Framework: The accurate diagnosis of FGR remains a significant limitation in clinical practice, as current techniques cannot readily distinguish constitutionally small, but physiologically normal fetuses, from pathologic growth restriction resulting from placental insufficiency. Moreover, the cohort we have does not differentiate between FGR fetuses presenting Doppler abnormalities and those without. The larger intra-classification variability in FGR may therefore represent a much larger spectrum than healthy controls alone. The three main characteristics of our data set were: (1) the over-representation of control MRI data when compared to FGR data, (2) the absence of any completely discriminating feature and (3) a higher clinical heterogeneity in the FGR group compared to controls. We thus developed a prediction framework based on RUSBoost (27) to obtain a set of weak learners to be used for classifying each subject as healthy or FGR. Specifically developed to deal with skewed data sets, RUSBoost is a variation of the Adaboost algorithm (28) that incorporates data sampling to deal with imbalanced learning sets (27). The RUSBoosting based framework was implemented in MATLAB®, with 1000 trees with a minimal depth of 5 as weak learners, and with surrogate splits. Prior probabilities were adjusted to penalize more false negatives than false positives and the FGR data were oversampled with a ratio 2 to 1.

Features selection: Given that the number of features is high with regard to the size of the database, diagnosis classification was performed on both the entire set of features as well as a selected subset. Two methods of features selection were investigated: (1) features ranking using perturbation experiments and (2) features selection using the ReliefF method (29).

To assess the impact of features selection, (the use of only a subset of the placental features), on the identification of FGR, the following procedure was used and repeated 50 times to account for sampling variability: using a 10-fold cross validation approach, the dataset was separated into a training set and a testing set. On the training set, a 10-fold cross validation was used to determine the optimal set of features. A cut-off at 5% was set: all attributes with a relative weight below 5% were considered non-essential. Finally, the best set of attributes, *i.e.*, the attributes used more than 50% of the time during the 10 iterations, were selected.

Birth Weight Estimation

Placental signature: Since we sought to predict weight at a given GA (namely the GA at birth), we needed a target GA from which to compute the prediction. Therefore, the GA at birth was added to the previously described placental signature. Given that the condition (healthy vs. FGR) does influence BW, condition was also added to the placental signature.

Framework: We used regression Forests (RgF) (30) as they are known for their ease of parameterization and their ability to deal with high dimensional data without overfitting. 1000 regression trees with surrogate splits and minimal depth of 5 were used. Since RgF are not prone to overfitting with the increase of dimensions (31), no features selection was investigated for BW estimation.

Validation

While the use of two scans per patient in a subset of both healthy and FGR cohorts may limit intra-cohort variability, in contrast to what is traditionally performed in statistical analysis, data modeling is not typically carried out using this type of machine learning approach (31). As such, inter- and intra-group differences between controls and FGR were not computed making the identification of FGR vs controls and BW estimation less sensitive to redundant data within each cohort. Moreover, the data in the testing set were blinded to the training set and all the experiments were performed per subject, *i.e.*, the same subject could belong to both the testing and the training set. A 10-fold cross validation approach repeated 50 times was performed to obtain all the results in this paper. For each repetition, learning and testing sets were built by selecting randomly the patients (rather than the MRI scans) ensuring that no data belonging to a given patient could be found simultaneously in both groups.

Manual Segmentation Validation—The dice coefficient

$\frac{2 |Segmentation_1 \cap Segmentation_2|}{(|Segmentation_1| + |Segmentation_2|)}$ was used to compute inter and intra-rater reproducibility on a set of 12 randomly selected placental volumes segmented three times; the first time by an experienced annotator and the latter two by a neonatologist blinded to the first segmentation.

Validation Metrics in Identifying FGR—The following metrics were used to assess the framework performances in identifying *in utero* FGR.

Let nb_{tp} , the number of true positives, *i.e.*, the number of FGR detected as such, nb_{tn} , the number of true negatives, *i.e.*, the number of controls detected as such, nb_{fp} , the number of false positives, nb_{fn} , the number of false negatives, nb_{fgr} , the number of FGR and $nb_{healthy}$, the number of controls.

1. Global accuracy (SR): $\frac{nb_{tp} + nb_{tn}}{nb_{fgr} + nb_{healthy}}$
2. Precision (P): $\frac{nb_{tp}}{nb_{tp} + nb_{fp}}$
3. Recall (R) or sensitivity: $\frac{nb_{tp}}{nb_{tp} + nb_{fn}}$
4. Specificity (SP): $\frac{nb_{tn}}{nb_{healthy}}$
5. Area Under the Receiver Operating Characteristic Curve (AUC):
 $\int_{-\infty}^{\infty} Recall(T) FPR'(T) dT$ with FPR (False Positive Rate) = $\frac{nb_{fp}}{nb_{healthy}}$
6. Diagnostic odds ratio (DOR): $\frac{recall * specificity}{(1 - recall) * (1 - specificity)}$

While the first three metrics are commonly used to evaluate prediction results, the unbalanced aspect of our dataset may hinder the correct interpretation. Therefore, we also used Receiver Operating Characteristic curves (32) to further validate our results (27).

To assess the choice of the classifier used and the impact of features selection, the proposed framework results were also compared to the ones obtained using RUSBoost with both methods of features selection, Support Vector Machines (SVM) (33) with and without features selection using ReliefF and Random Forests (30).

Validation Metrics in Predicting Birth Weight—To validate the proposed BW prediction method, the following metrics were used:

1. Mean Percentage Error (MPE): $\frac{1}{N} \sum_{i=1}^N \frac{BirthWeight - PredictedWeight}{BirthWeight} * 100$,
that computes the systematic deviation from the actual BW
2. Random Error (RE) or standard deviation of the error in percentage
 $\sqrt{\frac{1}{N-1} \sum_{i=1}^N \left| \frac{PredictedWeight - BirthWeight}{BirthWeight} - MPE * 0.01 \right|^2} * 100$
3. The proportions of predicted BW (PPB) within the 5%, 10% and 20% of the actual BW.

The results were compared to the ones obtained using SVM for regression (SVMr) (34) and least-squares boosting (LSBoost) (35 with and without features selection using RreliefF. Since our features all have different ranges, a normalization between [0,1] was performed for SVMr.

Binning Scheme Impact—To validate the choice of the number of bins used for the grey level reduction in the textural analysis, we performed a comparison of the results obtained using 16, 32, 64, 128 and 256 bins. An analysis of variance was performed to compare results at each grey level for each method followed by a multiple comparison with a Tukey-Kramer correction.

Predictors Ranking Method—To determine the relevance of each feature in both (1) identifying FGR and (2) estimating BW, we assessed the performance and importance of each feature. The contribution of each feature to either the classification (i.e., identification of FGR vs control) or the regression framework differs from direct correlation computation between the response variable and the tested feature. Rather, features ranking provides a strong indication of the validity of the use of a feature within the used framework. Therefore, the performance assessments were carried out by perturbation experiments before trees pruning since they were grown in both frameworks with surrogates.

RESULTS

Manual Segmentation Validation

We obtained respective Dice indices of 0.83 and 0.87 for intra-rater and an inter-rater reproducibility.

Identifying FGR

Table 3 presents the accuracy results of the proposed framework as well as the results obtained with other state-of-the-art classification methods as detailed in the method's section. The most robust results were obtained using RUSBoost with the full set of shape and textural features. This method distinguished healthy from FGR pregnancies with an accuracy of 86%, a precision of 77% and a recall of 86%. As an indicator for the high discriminatory abilities of the proposed method, the AUC rated at 0.86. Specificity was 87%, which lead to a Diagnostic Odd Ratio of 44.7, reflecting the high-discriminative abilities of the framework. Noteworthy, with features ranking using the perturbation experiment, none of the features had a weight amounting to more than 12% of the global weights. Only one feature had a weight superior to 10%, and the remainder had weights between 1% and 8% which likely explains the dramatic drop of performance observed when features pre-selection was performed. The same trends can be seen with the features selected by ReliefF as well.

The most important features found in FGR identification were the long run emphasis, the normalized mean grey levels, the long run high grey levels emphasis, the volume and the normalized kurtosis of the grey levels distribution. However, none of the predictors had a

strikingly high weight and thus none seems to be predominantly leading the diagnosis prediction.

Birth Weight Estimation for FGR and Healthy Fetuses

The results of BW estimation accuracy, computed as detailed in the method's section, are presented in Table 4. The average estimated BW for healthy fetuses using this machine-learning approach was 3088 grams $\pm$ 464, compared to the actual BW of 3366 grams $\pm$ 390. More specifically, the proposed framework predicted BW with mean percent error of 0.3% $\pm$ 13.4% (MPE $\pm$ RE); 32% of predictions were within 5% of actual BW, 59% within 10% of actual BW, and 90% were within 20% of actual BW. Within FGR, the estimated average BW was 2803 grams $\pm$ 580 compared to actual BW of 2345.6 grams $\pm$ 652 with a MPE of $-2.6 \pm 15.9\%$ (RE). The proportion of predicted BW within 5% of actual BW was 35%, 58% were within 10% of actual BW and 87% within 20% of actual birth weight. As with other BW estimation methods, there was a tendency to underestimate BW in FGR (mean error = -69 g) and in healthy controls (mean error = -9 g).

Within the context of BW estimation, the most important features were GA at birth, condition (i.e., control vs. FGR), long run high grey levels emphasis, placental volume and placental elongation. Of note, when reducing the placental signature to only GA at birth, BW prediction was $-1.9 \pm 13.3\%$ for controls and $0.7 \pm 22\%$ for FGR fetuses which confirms the need to develop dedicated frameworks for pregnancies complicated by FGR.

Choice of the Number of Bins Analysis

The impact of the grey level reduction on each framework was also investigated. For identifying FGR, the best results were obtained for the combination of RUSBoost with no features selection and a number of bins equal to 64 or higher. For RUSBoost with no features selection, results at 16 and 32 bins were significantly different from each other and from the results obtained at the other grey levels. There was no significant difference between the results obtained at 64, 128 and 256 levels. The number 128 was thus been chosen in the simulations on healthy vs. FGR diagnosis, given that it was the one optimizing the raw mean performance estimations. For BW estimation, the best results were obtained for RgF for all grey levels. There was no statistical difference between the results obtained with RgF for the different grey levels. The number 32 was thus chosen as it was the one optimizing the raw mean performance estimations. Of note, the most important features in estimating birth weight were the same for both 32 and 128 bins. (Supplementary figures S1 and S2 provide additional details).

DISCUSSION

We described a novel, semi-automated framework for identifying FGR and estimating BW using in vivo MRI placental features obtained in FGR and healthy fetuses during the second and third trimesters of pregnancy. We demonstrated a very high prediction performance in distinguishing FGR from healthy fetuses in utero, as well as a high prediction ability in estimating birth weight for both cohorts.

While existing studies have developed automated gestational age prediction tools based on brain features (36), there are, to the best of our knowledge, currently no available frameworks for the automated or semi-automated diagnosis of FGR or prediction of BW using in vivo placental features extracted from structural MRI images. No clinical method currently achieves adequate sensitivity and specificity in the detection of FGR. The abdominal circumference method plotted against the antenatal intergrowth 21st project chart achieves a specificity ranging from 94 – 96% but exhibits poor sensitivity of 35 – 41% (7). Interestingly, methods based on abdominal circumference plotted on the most widely used antenatal Hadlock charts achieve a sensitivity ranging from 74 – 82% and a specificity of 69 – 98% (7). The proposed framework outperforms these standard clinical tools by proposing a good trade-off in terms of sensitivity and specificity with a success rate of 86% for diagnostic accuracy, a sensitivity of 86% and a specificity of 87%.

Similarly, clinically available tools have various limitations in the estimation of birth weight, particularly for infants at the two ends of the spectrum of normal birth weight distribution (i.e., small for gestational age and large for gestational age infants). While the reasons are not well understood, it has been shown that most fetal weight estimation methods perform poorly when dealing with SGA fetuses (6). Previous studies have shown that methods specifically designed for SGA populations work better than generic ones (6). Moreover, even with SGA specific methods, those that consider SGA as a homogeneous group tend to overestimate BW while the opposite trend is observed for SGA sub-group specific approaches (6).

In this study, we investigated the usefulness of a placenta based MRI method to predict BW in FGR and healthy fetuses. Unlike previously described methods (7), this method, which also considers SGA as a homogeneous group, tended to underestimate BW for FGR fetuses. In our cohort, the inclusion criteria were designed to best identify FGR secondary to placental insufficiency; subjects with questionable dates of conception were excluded, but did include both fetuses with asymmetric growth and/or abnormal fetal Doppler studies. The estimation of BW in FGR through this method is in line with current tools available in routine clinical practice. Noteworthy, most comparative studies were done using measurements acquired within a week of delivery while in our study the mean GA at the time of the fetal MRI was 4.3 ± 6.1 weeks prior to delivery for FGR fetuses. For healthy fetuses our results for birth weight estimation slightly outperformed what is currently used in clinical practice (37) using MRI measures that were obtained up to 8 weeks (7.9 ± 5 weeks) before delivery. Our method was thus shown to be reliable throughout the second half of gestation, which allowed for earlier detection of fetuses at risk for intrauterine growth restriction. The early, accurate detection of fetal growth impairment is critical to prevent fetal compromise. Indeed, early biomarkers of placental disease that may impact fetal growth and well-being open a new realm of possibilities in terms of *in utero* intervention to augment placental function that may ameliorate or even reverse fetal growth delays.

Over the last few years there have been notable advances in deep learning methods that have resulted in automated generation of features. We elected to confine our analysis to a set of hand-crafted ones given that purely automated methods often require the use of large datasets to optimize learning. Moreover, the main aim of this work was to investigate the

placenta directly in relation to fetal growth. As such, we wanted to be able to directly link explicit shape and textural features of the in vivo placenta to fetal outcome, namely the identification of FGR and prediction of birth weight. Interestingly, the importance of each feature differed for identifying FGR and predicting BW. Collectively, these findings highlight that future studies are urgently needed to understand the underlying placental mechanisms driving in utero fetal growth disturbances.

Both fetal growth and neonatal birth weight are influenced by numerous recognized and unrecognized genetic, environment and physiologic factors, including parental race, stature, nutrition, and altitude, even in the absence of overt pathology (38). In this report, we provide a methodology to interrogate intrinsic placental features on pregnancy outcomes, independent of known and unknown clinical risk factors. Nonetheless, future studies are warranted to combine this method with additional clinical characteristics to best reflect the multi-factorial determinants of fetal weight.

Although our paper has a number of strengths, the limitations deserve mention. From a methodological point of view, one of the primary limitations of our study is the use of a manual placental segmentation. Automation of placental segmentation on MRI data remains a challenging task due to its variability in shape, orientation, position and appearance (39). To the best of our knowledge, only two methods currently exist tackling the problem of automatic (39) or semi-automatic placenta segmentation on MR data (40). In (39), the authors reported an accuracy of $71.95 \pm 19.79\%$ for healthy fetuses and $66.89 \pm 15.35\%$ for a cohort mixing FGR fetuses. The semi-automated segmentation proposed in (40) presented better results on healthy fetuses taking advantage of multiple views of the same placenta. However, our study protocol only included a single orientation acquisition; therefore we felt manual segmentations would provide the most optimal results. Noteworthy, the inter- and intra-rater segmentations were reasonably high, underscoring the reliability and reproducibility of our manual placental measures. Another limitation of our study concerns sample size. Indeed, the use of two scans for a subset of the fetuses may have reduced the variability of the data and the subsequent results. As such, this study needs to be confirmed on a larger cohort. The limited sample size may also have affected the sensitivity rate in distinguishing FGR from controls, as well as impacted the variability in estimating birth weight for both populations. Moreover, our cohort does not differentiate between FGR fetuses with Doppler changes and those without. The larger intra-classification variability in FGR may therefore represent a much larger spectrum of disease than in healthy controls alone. Our dataset consists of both axial and coronal acquisitions and the distinction between both was not factored in the analysis, which could be constructed as a limitation of our study. However we were underpowered to examine the potential effect of acquisition plane on our prediction results. Finally, while inter-slice motion was not explicitly taken into account in the proposed method, textural features were computed and averaged over the whole placenta minimizing the impact of motion on the computed features.

In conclusion, we propose a novel method to (1) to identify fetal growth failure in utero by improving the diagnostic accuracy of FGR in our cohort and (2) estimating birth weight across both our healthy and high risk populations using a combination of placental shape and textural features computed on 3D MR images. Ongoing work is necessary to investigate

the different predictors influencing feto-placental development in other high-risk populations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank our clinical research coordinating team and the families that participated in this study.

Grant Support: This publication was supported by Award Numbers CTSI-UL1TR000075, CTSI-KL2 TR000076 from the Clinical and Translational Science Institute (Children's National Health System) and NHLBI R01 HL116585-04 from the National Heart, Lung, and Blood Institute (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Heart, Lung, and Blood Institute or the Clinical and Translational Science Institute.

References

1. Gude NM, Roberts CT, Kalionis B, King RG. Growth and function of the normal human placenta. *Thrombosis Research*. 2004; 114:397–407. [PubMed: 15507270]
2. Kim CJ, Romero R, Chaemsathong P, Kim JS. Chronic inflammation of the placenta: definition, classification, pathogenesis, and clinical significance. *American Journal of Obstetrics and Gynecology*. 2015; 213:S53–S69. [PubMed: 26428503]
3. Gagnon R. Placental insufficiency and its consequences. *European Journal of Obstetrics & Gynecology and Reproductive Biology*. 2003; (Supplement, 110):S99–S107. [PubMed: 12965097]
4. Lindqvist PG, Molin J. Does antenatal identification of small-for-gestational age fetuses significantly improve their outcome? *Ultrasound in Obstetrics and Gynecology*. 2005; 25:258–264. [PubMed: 15717289]
5. Bardien N, Whitehead CL, Tong S, Ugoni A, McDonald S, Walker SP. Placental insufficiency in fetuses that slow in growth but are born appropriate for gestational age: A prospective longitudinal study. *PLoS ONE*. 2016; 11:1–13.
6. Melamed N, Ryan G, Windrim R, Toi A, Kingdom J. Choice of formula and accuracy of fetal weight estimation in small-for-gestational-age fetuses. *Journal of Ultrasound in Medicine*. 2016; 35:71–82. [PubMed: 26635253]
7. Poljak B, Agarwal U, Jackson R, Alfirevic Z, Sharp A. Diagnostic accuracy of individual antenatal tools for the detection of the small for gestational age newborn. *Ultrasound in Obstetrics & Gynecology*. 2017; 49:493–499. [PubMed: 27486031]
8. Schwartz N, Wang E, Parry S. Two-dimensional sonographic placental measurements in the prediction of small-for-gestational age infants. *Ultrasound in Obstetrics & Gynecology*. 2012; 40:674–9. [PubMed: 22331557]
9. Plasencia W, Akolekar R, Dagklis T, Veduta A, Nicolaides K. Placental volume at 11–13 weeks' gestation in the prediction of birth weight percentile. *Fetal Diagnosis and Therapy*. 2011; 30:23–8. [PubMed: 21701132]
10. Fisteag-Kiprono L, Neiger R, Sonek JD, Croom CS, McKenna DS, Ventolini G. Perinatal outcome associated with sonographically detected globular placenta. *Journal of Reproductive Medicine*. 2006; 51:563–566. [PubMed: 16913547]
11. Chen KH, Chen LR, Lee YH. Exploring the relationship between preterm placental calcification and adverse maternal and fetal outcome. *Ultrasound in Obstetrics & Gynecology*. 2011; 37:328–34. [PubMed: 20586039]
12. Ohgiya Y, Nobusawa H, Seino N, et al. MR Imaging of Fetuses to Evaluate Placental Insufficiency. *Magnetic Resonance in Medical Sciences*. 2016; 15:212–9. [PubMed: 26607809]
13. Damodaram M, Story L, Eixarch E, et al. Placental MRI in intrauterine fetal growth restriction. *Placenta*. 2010; 31:491–498. [PubMed: 20347139]

14. Hadlock FP, Deter RL, Harrist RB, Park SK. Estimating fetal age: computer-assisted analysis of multiple fetal growth parameters. *Radiology*. 1984; 152:497–501. [PubMed: 6739822]
15. Alliez P, Jamin C, Mérigot Q, et al. Point set processing. *CGAL User and Reference Manual*, CGAL Editorial Board. 2016
16. Kazhdan M, Bolitho M, Hoppe H. Poisson surface reconstruction. *Proceedings of the Fourth Eurographics Symposium on Geometry Processing*. 2006:61–70.
17. Whittle W. Ultrasound detection of placental insufficiency in women with ‘unexplained’ abnormal maternal serum screening results. *Clinical Genetics*. 2006; 69:97–104. [PubMed: 16433688]
18. Lee T, Kashyap R, Chu C. Building skeleton models via 3-D medial surface axis thinning algorithms. *CVGIP: Graphical Models and Image Processing*. 1994; 56:462–478.
19. Kerschnitzki M, Kollmannsberger P, Burghammer M, Duda GN, Weinkamer R, Wagermaier W, Fratzl P. Architecture of the osteocyte network correlates with bone material quality. *Journal of Bone and Mineral Research*. 2013; 28:1837–1845. [PubMed: 23494896]
20. Shapira L, Shamir A, Cohen-Or D. Consistent mesh partitioning and skeletonisation using the shape diameter function. *Visual Computer*. 2008; 24:249–25.
21. Haralick RM. Statistical and structural approaches to texture. *Proceedings of the IEEE*. 1979; 67:786–804.
22. Connors RW, Harlow CA. A theoretical comparison of texture algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1980; 2:204–222. [PubMed: 21868894]
23. Galloway MM. Texture analysis using gray level run lengths. *Computer Graphics and Image Processing*. 1975; 4:172–179.
24. Dasarathy BV, Holder EB. Image characterizations based on joint gray level-run length distributions. *Pattern Recognition Letters*. 1991; 12:497–502.
25. Chu A, Sehgal CM, Greenleaf JF. Use of gray value distribution of run lengths for texture analysis. *Pattern Recognition Letters*. 1990; 11:415–419.
26. Xu DH, Kurani AS, Furst JD, Raicu DS. Run-Length Encoding for Volumetric Texture. *Proceedings of the Iasted international conference on visualization imaging and image processing*. 2004
27. Seiffert C, Khoshgoftaar TM, Hulse JV, Napolitano A. Rusboost: A hybrid approach to alleviating class imbalance. *IEEE Transactions on Systems, Man, and Cybernetics*. 2010; 40:185–197.
28. Freund Y, Schapire RE. Experiments with a new boosting algorithm. *International Conference on Machine Learning*. 1996:148–156.
29. Robnik-Sikonja M, Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning*. 2003; 53:23–69.
30. Breiman L. Random forests. *Machine Learning*. 2001; 45:5–32.
31. Breiman L. Statistical Modeling: The Two Cultures. *Statistical Science*. 2001; 16:199–231.
32. Provost FJ, Fawcett T. Robust classification for imprecise environments. *CoRR*. 2000
33. Vapnik V. Estimation of Dependences Based on Empirical Data. 1979
34. Drucker H, Burges CJC, Kaufman L, Smola A, Vapnik V. Support Vector Regression Machines. *Advances in Neural Information Processing Systems*. 1997; 9:155–161.
35. Friedman JH. Greedy Function Approximation: A Gradient Boosting Machine. 1999
36. Namburete AI, Stebbing RV, Kemp B, Yaqub M, Papageorgiou AT, Noble AJ. Learning-based prediction of gestational age from ultrasound images of the fetal brain. *Medical Image Analysis*. 2015; 21:72–86. [PubMed: 25624045]
37. Kurmanavicius J, Burkhardt T, Wisser J, Huch R. Ultrasonographic fetal weight estimation: accuracy of formulas and accuracy of examiners by birth weight from 500 to 5000 g. *Journal of Perinatal Medicine*. 2004; 32:155–61. [PubMed: 15085892]
38. Catalano PM, Drago NM, Amini SB. Factors affecting fetal growth and body composition. *American Journal of Obstetrics and Gynecology*. 1995; 172:1459–1463. [PubMed: 7755054]
39. Alansary A, Kamnitsas K, Davidson A, et al. Fast Fully Automatic Segmentation of the Human Placenta from Motion Corrupted MRI. *Medical Image Computing and Computer-Assisted Intervention*. 2016:589–597.

40. Wang G, Zuluaga MA, Pratt R, et al. Slic-Seg: A minimally interactive segmentation of the placenta from sparse and motion-corrupted fetal MRI in multiple views. *Medical Image Analysis*. 2016; 34:137–147. [PubMed: 27179367]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.

(a) Axial slice of a 3D MR placental image, (b) placental manual segmentation of the same axial slice, (c) placental thickness, computed using SDF on the 3D mesh, represented as a heat map (normalized between 0 and 1 for visualization purposes) and (d) skeleton of the placenta (in red) from which elongation was derived.

Table 1

Baseline characteristics of our cohort.

	Controls N=46	FGR N=34	p-value
GA at MRI 1	28 ± 3.7	31 ± 3.8	<0.05
GA at MRI 2	35.9 ± 2.1	35.8 ± 0.5	<0.05
GA interval	[18.8 – 39.7]	[22.4 – 37]	0.8
GA at birth	39.6 ± 1.2	36.3 ± 6.47	<0.05
GA at MRI to delivery interval (weeks)	7.9 ± 5.1	4.3 ± 6.1	<0.05
% Female fetuses	50%	61.7%	0.08
Unknown gender fetuses	2	1	
Birth weight (grams)	3366 ± 390	2345.6 ± 652	<0.05
Birth weight (z-score)	– 0.07 ± 0.62	– 1.5 ± 0.76	<0.05
Number of fetuses	46	34	
Number of single scans	10	26	
Number of double scans	36	8	
Total number of scans	82	42	

Table 2

Description of the textural features based on the GLCM and Run-length matrices used in this study (21 – 26).

Feature name	Formula	Interpretation
<p>Let Ω, the image to be studied, p the normalized grey level co-occurrence matrix (GLCM), i and j two grey levels and $p(i,j)$ the element at i, j of the GLCM p and $\mu_i = \sum_{i,j} i \cdot p(i,j)$ and $\mu_j = \sum_{i,j} j \cdot p(i,j)$</p>		
Energy	$\sum_{i,j} p(i,j)^2$	uniformity measurement
Entropy	$-\sum_{i,j} p(i,j) \log_2 p(i,j), \text{ if } p(i,j) > 0$	measure of randomness
Inverse Difference Moment	$\sum_{i,j} \left(\frac{1}{1+(i-j)^2} \right) \cdot p(i,j)$	measure of local homogeneity
Contrast	$\sum_{i,j} (i-j)^2 \cdot p(i,j)$	measure of the local intensity variation
Cluster Shade	$\sum_{i,j} ((i - \mu_i) + (j - \mu_j))^3 \cdot p(i,j)$	lack of symmetry measurement
Cluster Prominence	$\sum_{i,j} ((i - \mu_i) + (j - \mu_j))^4 \cdot p(i,j)$	lack of symmetry measurement
<p>Let R be the number of runs and p the run-length matrix defining the number of occurrence of pixels of grey level i (that have been quantized into G grey levels) at run length j and direction Θ.</p>		
Short Run Emphasis	$\frac{1}{R} \sum_{i,j} \left(\frac{p(i,j)}{j^2} \right)$	measure of short runs distribution: higher when the number of short runs is higher than the number of long runs i.e., in finer textures
Long Run Emphasis	$\frac{1}{R} \sum_{i,j} \left(p(i,j, \Theta) * j^2 \right)$	complementary metric, i.e., higher in coarse textures or with large homogeneous areas
Grey Level Non uniformity	$\frac{1}{R} \sum_{i,j} (p(i,j, \Theta))^2$	increases with the number of outliers
Run Length Non Uniformity	$\frac{1}{R} \sum_{j,i} (p(i,j, \Theta))^2$	examines the distribution of run lengths, higher when the texture is dominated by a few run lengths outliers
Low Grey Level Run Emphasis	$\frac{1}{R} \sum_{i,j} \left(\frac{p(i,j, \Theta)}{i^2} \right)$	determines the distribution of runs of low grey level, higher when the texture is dominated by the low grey levels runs
High Grey Level Run Emphasis	$\frac{1}{R} \sum_{i,j} \left(p(i,j, \Theta) \cdot i^2 \right)$	complementary metric to the previous one

Feature name	Formula	Interpretation
Short Run Low Grey Level Emphasis	$\frac{1}{R} \sum_{i,j} \left(\frac{p(i, j, \Theta)}{i^2 \cdot j^2} \right)$	higher when there is a high number of short runs of low grey levels
Short Run High Grey Level Emphasis	$\frac{1}{R} \sum_{i,j} \left(\frac{p(i, j, \Theta) \cdot i^2}{j^2} \right)$	complementary metric to the previous one for high grey levels
Long Run Low Grey Level Emphasis	$\frac{1}{R} \sum_{i,j} \left(\frac{p(i, j, \Theta) \cdot j^2}{i^2} \right)$	higher when there is a high number of long runs of low grey levels
Long Run High Grey Level Emphasis	$\frac{1}{R} \sum_{i,j} \left(p(i, j, \Theta) \cdot i^2 \cdot j^2 \right)$	complementary metric to the previous one for high grey levels

Table 3

Comparison of diagnostic accuracy with regard to the classifier.

Algorithm/Measure	Accuracy $\pm$ std %	Precision $\pm$ std %	Recall $\pm$ std %	AUC $\pm$ std	Specificity $\pm$ std %	DOR $\pm$ std
RUSBoost	86 $\pm$ 2	77 $\pm$ 3	86 $\pm$ 3	86 $\pm$ 2	87 $\pm$ 2	44.7 $\pm$ 15
RUSBoost-perturbation	65 $\pm$ 3	51 $\pm$ 3	95 $\pm$ 2	72 $\pm$ 2.8	50 $\pm$ 4.6	22 $\pm$ 10
RUSBoost-Relief	66 $\pm$ 2	50 $\pm$ 1.8	94 $\pm$ 2	73 $\pm$ 2	52 $\pm$ 3	19.5 $\pm$ 9
SVM	80 $\pm$ 1	75 $\pm$ 2	63 $\pm$ 3.6	76 $\pm$ 1.8	89 $\pm$ 1	14.4 $\pm$ 2.6
SVM-Relief	72 $\pm$ 2	71 $\pm$ 7.7	29 $\pm$ 5	62 $\pm$ 2.8	94 $\pm$ 1.7	7.49 $\pm$ 4.7
RF	83 $\pm$ 1.8	76 $\pm$ 2.7	70 $\pm$ 4.4	80 $\pm$ 2	89 $\pm$ 1	22.6 $\pm$ 5

Table 4

Comparison of BW prediction accuracy measured using MPE $\pm$ RE, PPB and BW mean error.

	Healthy fetuses			FGR fetuses		
	MPE $\pm$ RE	PPB [5 10 20]%	Mean error	MPE $\pm$ RE	PPB [5 10 20]%	Mean error
Rgf	0.3 $\pm$ 13.4	[32 59 90]	-69g	-2.6 $\pm$ 15.9	[35.5 58 87]	-9g
LSBoost	-0.6 $\pm$ 16	[27 48 78]	36g	-1.3 $\pm$ 18	[28 46 73]	41g
LSBoost-perturbation	14.8 $\pm$ 14.3	[25 51 84]	34g	-6.2 $\pm$ 26.7	[20 41 69]	-48g
LSBoost-Rreleff	-0.16 $\pm$ 15.3	[25 48 82]	24g	-9 $\pm$ 38	[19 37 63]	-55g
SVMr	7 $\pm$ 12	[23 41 83]	284g	-48.8 $\pm$ 74	[10 25 42]	-728g
SVMr-Rreleff	6.6 $\pm$ 12	[27 42 86]	268g	-48 $\pm$ 73	[9 26 43]	-720g