Predicting Birth Weights from Placental Surfaces Using a High-Dimensional Shape Descriptor

Jen-Mei Chang, Ph.D.[1], Amy Mulgrew[1], and Carolyn Salafia, M.D.[2]

[1]California State University, Long Beach, CA, U.S.A., [2]Placental Analytics, LLC, Larchmont, NY, U.S.A.

Overview

- This project seeks to understand the relationships between the health of a placenta and the baby.
- Previous studies have shown that the median placental shape at term is round, and deviation from such shape is related to a decreased placental functional efficiency.
- We propose the use of a nearly-continuous shape descriptor termed signed deviation vector to systematically capture the relationship between various maternal and fetal characteristics and the shape of the placental surface.
- Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) are used to produce classification statistics.
- The initial findings indicate significant relationships between shape of the placental surface and newborn's birth weight as well as their gestational age.

Why Study Placenta

- Recent medical research indicates that the placenta may be the crystal ball for the health of the baby.
- The placenta is the source of nutrition, oxygen, and blood for the developing fetus so any problem with the placenta may become a problem for the baby.
- An analysis of the placenta may help to predict risks for certain diseases that develop in the womb such as diabetes, autism, and heart disease.
- In particular, the structure of the blood vessel network as well as the shape of the human placenta may contain important medical clues.

Question: Which one of following placentas is associated with a healthy baby?

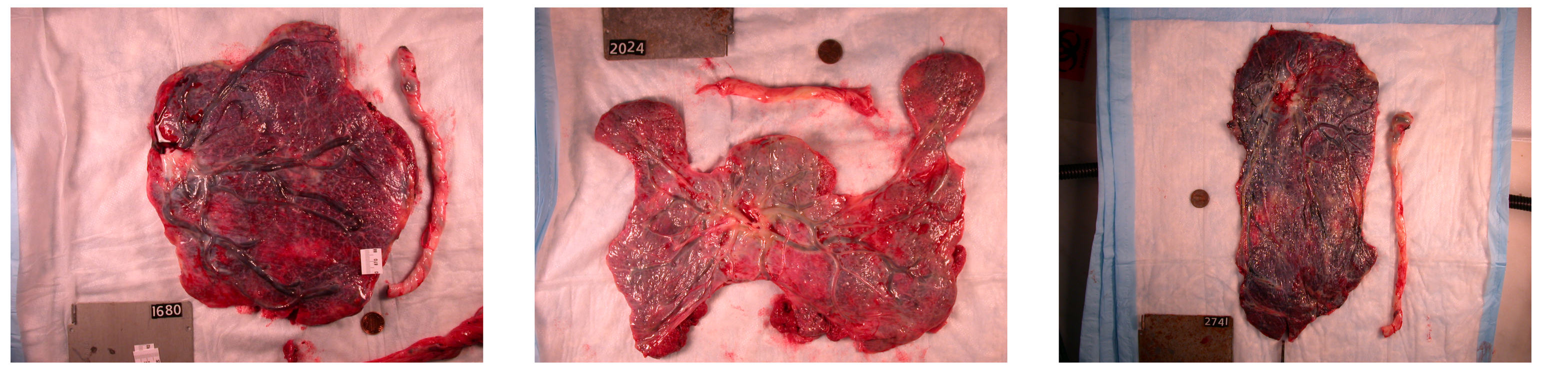


Figure: Sample digital placenta images in the UNC data set provided by Placental Analytics.

Preprocessing: Create Ground Truths

- **Hand Trace.** Of the placentas imaged, 150 were traced by hand by a trained pathologist. Using a Tablet PC and GNU Image Manipulation Program (GIMP), the perimeter was traced in green, the umbilical cord insertion marked in yellow and the blood vessels traced in pink.
- **Birth Weight Label.** 50 images were chosen at random from each group of low (BW < 2500 gram), normal (2500 gram $\leq$ BW ≤ 3500 gram), and high (BW > 3500 gram) birth weights.
- **Data Set Reduction.** Four of the tracings were poor, one was a duplicate, four were missing placenta weights and two were missing a maternal vascular pathology diagnosis. We arrive at 139 total placenta images for investigation.

⇓ Hand Traced

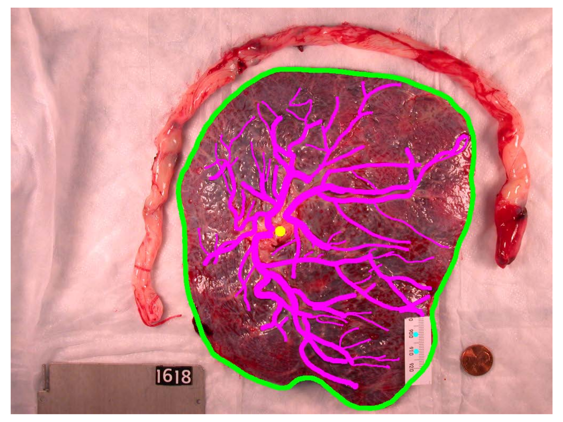


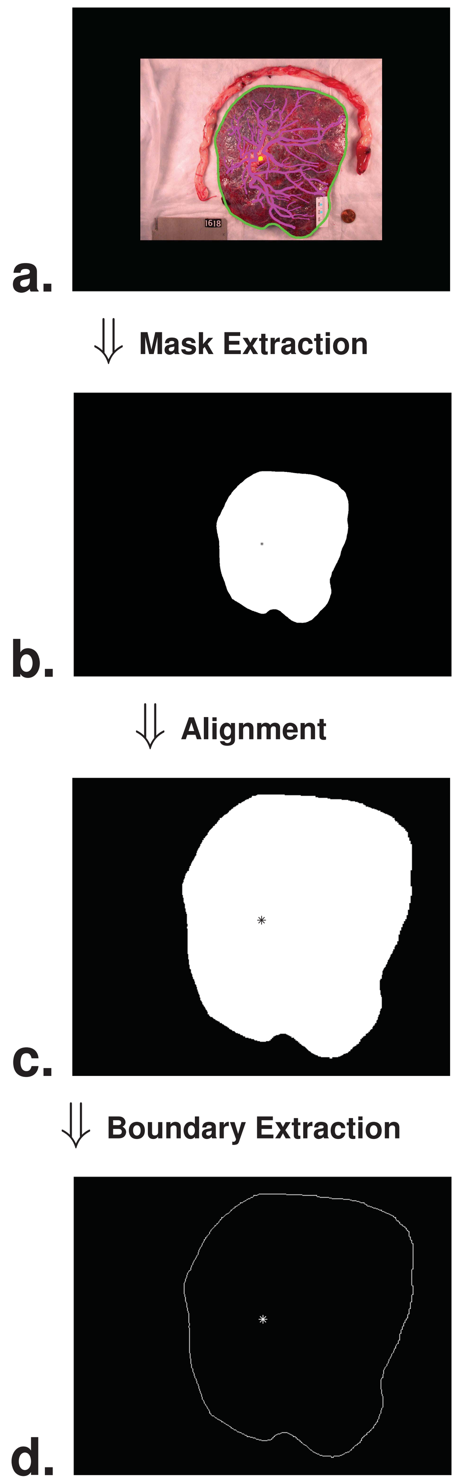
Image Registration

a. **Size Normalization.** All images were first normalized to size 1600-by-1200 for the ease of future computations.

b. **Mask Extraction.** The boundary of the placenta was extracted from the green perimeter hand tracing and filled with white to form the placental mask. The mean of the yellow pixel locations provided a single point for the umbilical insertion.

c. **Image Alignment.** The images are scaled to two pixels/mm and translated such that the umbilical insertion point rested in the center of each image. The images were further reduced to 797-by-1049 to reduce computational complexity.

d. **Boundary Extraction.** The boundary image, still of size 797-by-1049, of each placenta is obtained by eroding the boundary pixels into one pixel in width.



Shape Feature Extraction

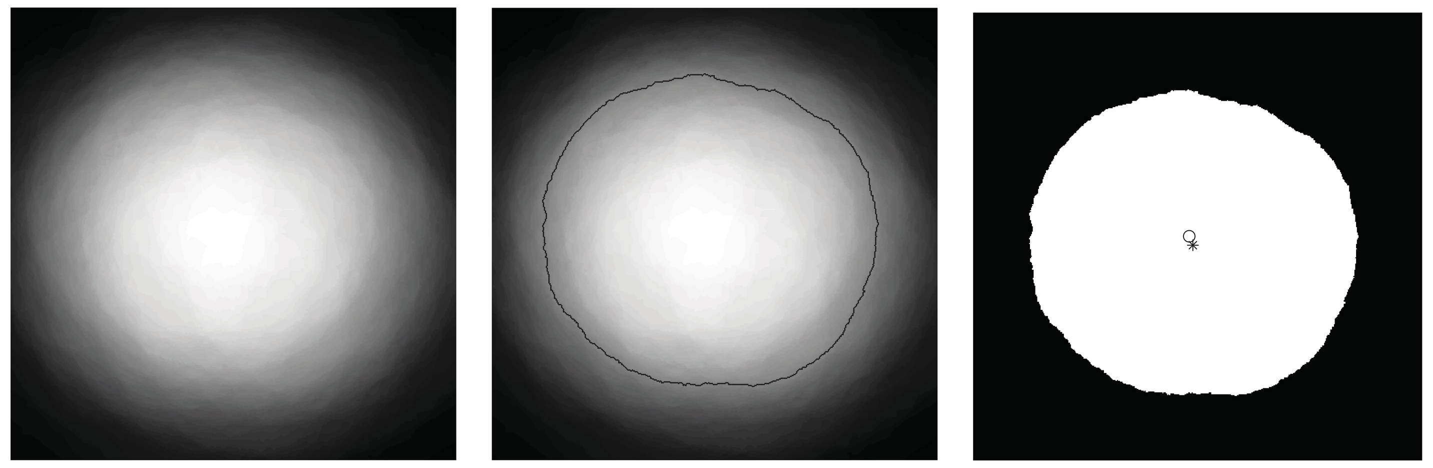
Shape of Normality

Previous studies have shown that an average placenta is round with the umbilical cord inserted in the center. Deviation from this prototypical placenta shape is related to a decrease in placental functional efficiency.

Let N be the total number of placenta images and $p^{(k)}$ a mask matrix, $1 \leq k \leq N$.

Define the mask frequency matrix $\mathcal{F}_{ij} = \sum_{k=1}^{N} p_{ij}^{(k)}$. Then the median placenta mask, $\mathcal{M}$, is created based on a majority rule:

$$\mathcal{M}_{ij} = \begin{cases} 0 & \text{if} \quad \mathcal{F}_{ij} < \frac{N}{2} \\ 1 & \text{if} \quad \mathcal{F}_{ij} \geq \frac{N}{2} \end{cases}$$

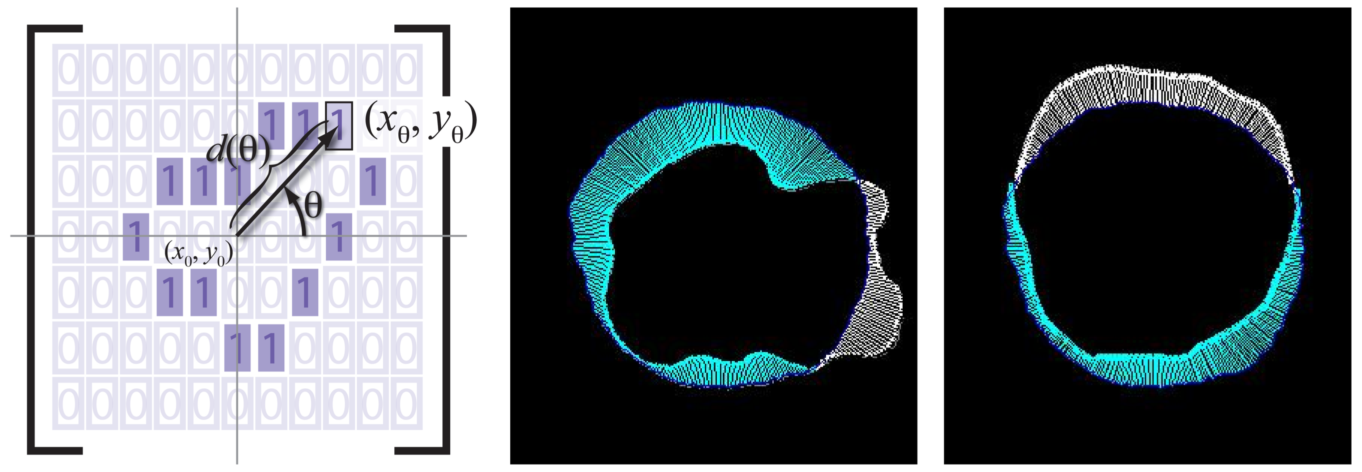


Frequency Matrix | Separation | Median Mask

Shape Feature Extraction

Signed Deviation Vectors (SDV)

- For each placenta, $P^{(i)}$, let $B^{(i)}$ be its boundary matrix (a pixel location is one if boundary exists, zero otherwise). A radial distance, $d(\theta)$, is defined to be the distance, in pixels, from the center of the image, (x_0, y_0), to the radial coordinate, (x_θ, y_θ), where θ is measured counterclockwise from the $+x$-axis.
- If $\mathbf{d}^{(i)} = [d(\theta)]_{\theta=1}^{k}$ ($k = 360$ in our studies), then the SDV for $P^{(i)}$ is defined as the unique expression $\mathbf{v}^{(i)} = \mathbf{d}^{(i)} - \bar{\mathbf{d}}$, where $\bar{\mathbf{d}}$ is the radial distance vector for the median placenta.



Radial Distance | SDV of LBW (1600g) | SDV of NBW (2761g)

High-Dimensional Shape Descriptor

- Principal Component Analysis (PCA) is performed on the collection of SDVs. And the projected coefficients, $\mathbf{PC}_i^{(\mu)}$, of the μ-th subject in the i-th principal direction, are used to capture shape features of each subject.

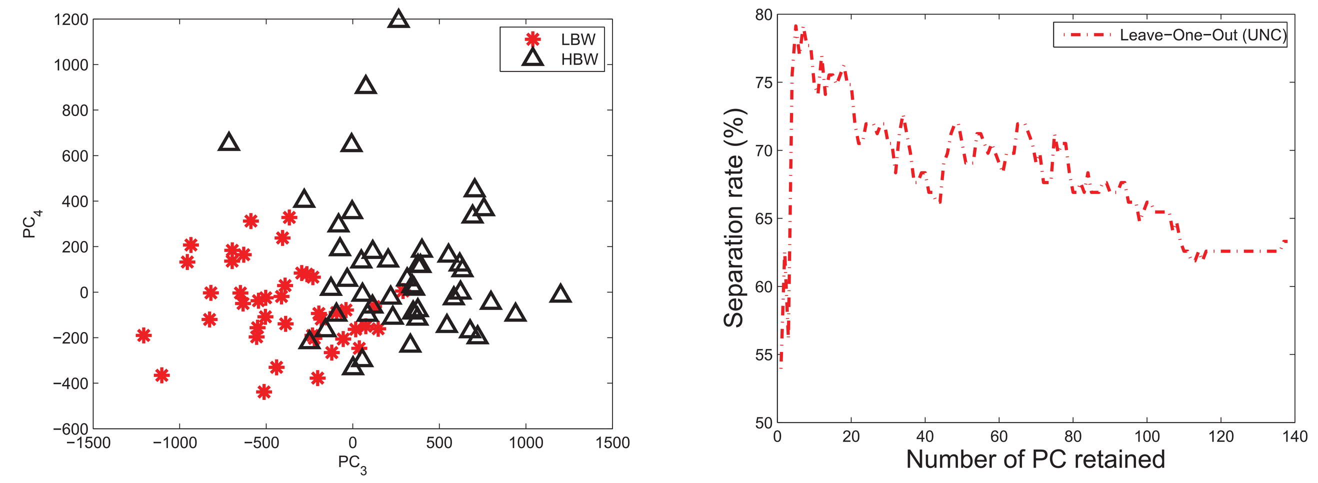


Figure: Left: Placental surface shapes of low-birth-weight babies are distinctively different than placental surface shapes of high-birth-weight babies. Right: The high separation rate is obtained with 3 **PC**s.

Method and Results

- A Leave-One-Out Cross Validation (LOOCV) is implemented with Linear Discriminant Analysis (LDA) to produce classifier statistics.
- $r = \max\left\{\frac{\sum_{i=1}^{N}\phi(\mathbf{x}_i)}{N}, 1 - \frac{\sum_{i=1}^{N}\phi(\mathbf{x}_i)}{N}\right\}$ gives the database's prior statistics, where $\phi(\mathbf{x}_i) = 0$ or 1. If the probability of the classifier with accuracy $0 \leq \mathsf{S} \leq 1$ produces a non-reliable classification outcome is given by $p(\mathsf{S}) = \sum_{k=\mathsf{S}N}^{N} \binom{N}{k} r^k \cdot (1-r)^{N-k}$, where $\mathsf{S}N$ gives the number of cases that are correctly identified, then the Classifier Confidence Rate (CCR) is defined as $\min_{0 \leq \mathsf{S} \leq 1} p(\mathsf{S}) < \epsilon$ (to ensure a confidence level of $1 - \epsilon$) and is used to gauge the statistical validity of the classifier.

Feature Label	$\frac{n_1}{N}$	$\frac{n_0}{N}$	CCR	S
BWT	30.22	69.78	76.98	82.73
Gender	61.15	38.85	68.35	60.43
Preterm vs. Term	23.08	76.92	83.45	83.45

Table: $\frac{n_1}{N}$ and $\frac{n_0}{N}$ give the percentage of the label 1 and label 0 group in the data set, respectively. CCR = classifier confidence rate with 95% confidence level, S = the best separation rate.

Department of Mathematics and Statistics, California State University Long Beach, Long Beach, CA 90840

jen-mei.chang@csulb.edu