Discriminating Placentas of Increased Risk for Autism with Chorionic Surface Vascular Network Features

Jen-Mei Chang, PhD

Department of Mathematics and Statistics
California State University, Long Beach
jen-mei.chang@csulb.edu

AWM Research Symposium, UCLA
April 8, 2017

Acknowledgement

The work presented here is the weighted $\sum$ of many people's contribution.

Statistics

- *Hui Zeng*, CSU Long Beach (graduate student); *Ruxu Han*, CSU Long Beach (graduate student); *Ya-Mei Chang*, Tamkang University

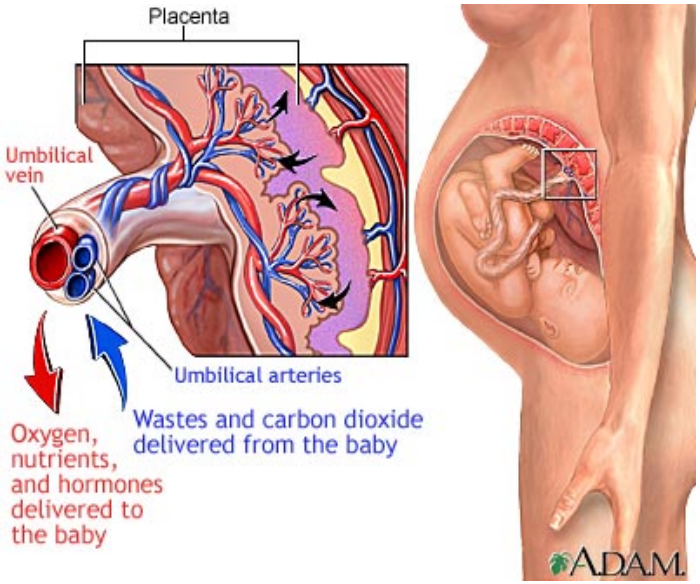
Data & research

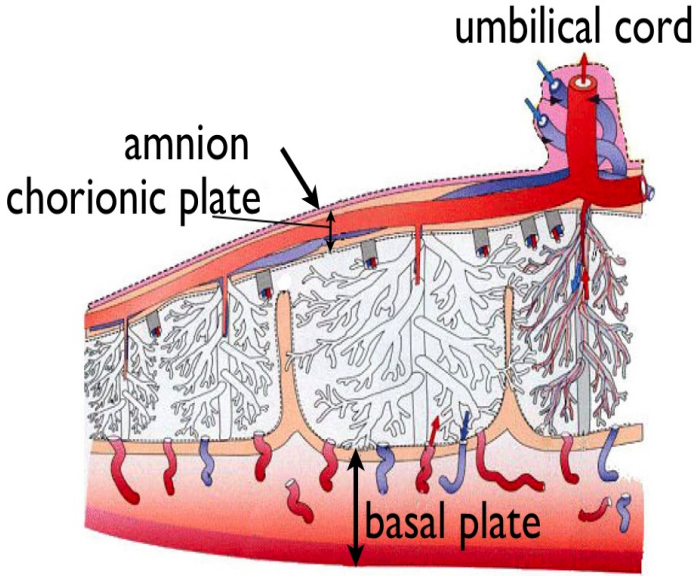
- *Carolyn Salafia*, Placental Analytics, LLC, Institute for Basic Research; *Ruchit Shah*, Placental Analytics, LLC

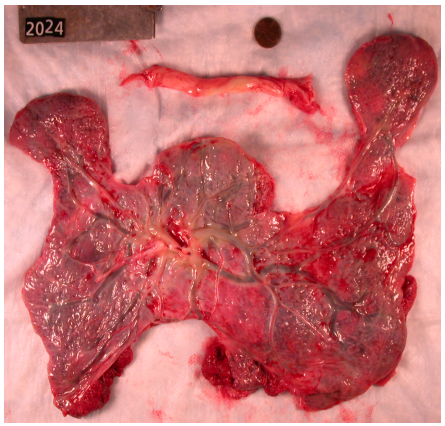
Data & medical content

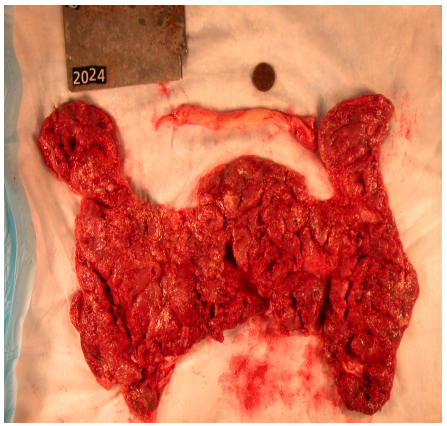
- *Craig Newschaffer*, Drexel University; *Richard Miller*, University of Rochester, NIH National Children's Study Placenta Consortium; *Philip J. Katzman*, University of Rochester, NIH National Children's Study Placenta Consortium; Jack Moye, NICHD; Margaret Fallin, Johns Hopkins University; Cheryl K. Walker, UC Davis; Lisa Croen, Kaiser Permenante Division of Research

Placenta & placental chorionic surface vascular network (PCSVN)









Autism spectrum disorder (ASD)

- First identified in 1943 by Leo Kanner.
- The May 2013 publication of the DSM-5 diagnostic manual merged all autism disorders into one umbrella diagnosis of ASD.
- Autism is a neurodevelopmental disorder with 3 defining areas of deficit:
  1. Social reciprocity
  2. Communication
  3. Restricted, repetitive patterns of behaviors, interests, or activities.
- Symptoms are developed by 36 months of age.

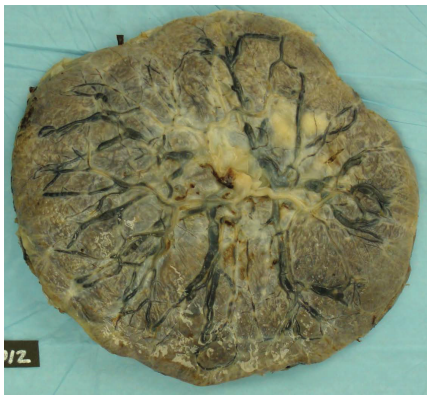
Placenta as a proxy for diagnosing ASD

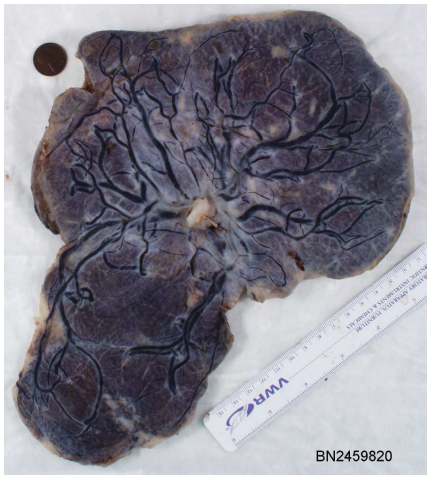
- There is not yet a proven cure for ASD.
- The brain is most responsive to treatment in the first year of life.
- Early intervention is associated with normalized brain activities in young children with ASD (e.g., improved social and communication skills.) [1].
- A diagnose of ASD is usually not made until the child is 3 or 4 years old, **the best opportunities for intervention have already been lost**.

- Altered patterns of angiogenesis $\Leftrightarrow$ variation in mature vascular network structures $\Leftrightarrow$ functional alterations of many viscera (e.g., lung, kidney, and pancreas).
- The gene families that control branching morphogenesis are shared between those permanent viscera and the temporary fetal organ – placenta.
- Placenta, hence, provides unique insights into the effects of genes and environment on key mechanisms required for conceptus development, including fetal origins of disease from hypertension to autism.

Research questions [2]

RQ: Which placenta is associated with an increased ASD risk?





RQ: Are there PCSVN features that distinguish placentas of increased risk for ASD from those in the general population? If so, what are they?

We are interested in discovering a reliable biomarker in assessing prenatal/neonatal ASD risk by considering the gestational origin of life – placenta.

Digital images of EARLI & NCS

- **89 EARLI placentas** (**increased risk** for ASD)

  Early Autism Risk Longitudinal Investigation (EARLI) [3] is an autism enriched-risk pregnancy cohort that focuses on the prenatal and early life periods of children who have biological siblings already diagnosed with ASD.

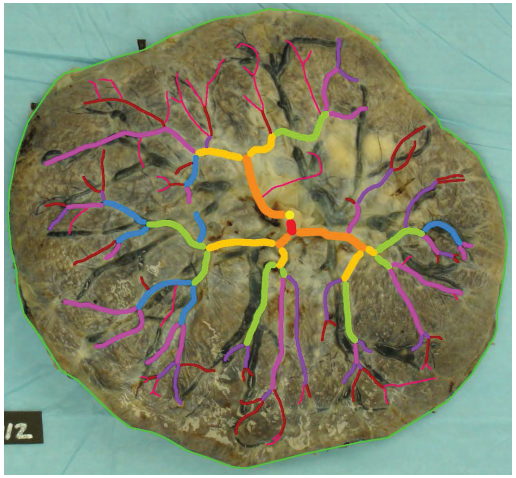
- **201 NCS placentas** (**normal risk** for ASD)

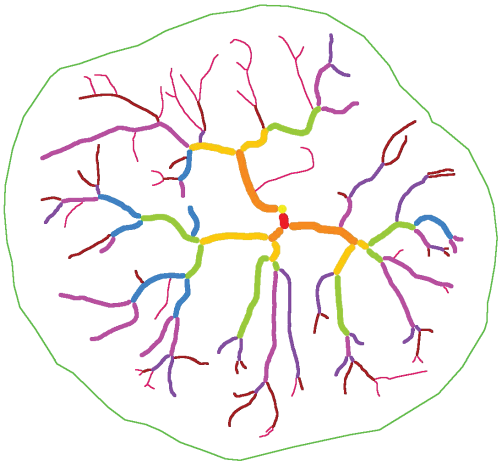
  National Children's Study (NCS) is a population-based cohort with pregnancies at unknown risk for ASD. NCS was designed to study environmental influences on child health and development and it enlisted participants without a bias towards risks and diagnoses in autism.

Digital photographs of the fetal surface were obtained on NCS & EARLI placentas following the same imaging protocol. The photos were taken either at delivery or upon pathology evaluation.

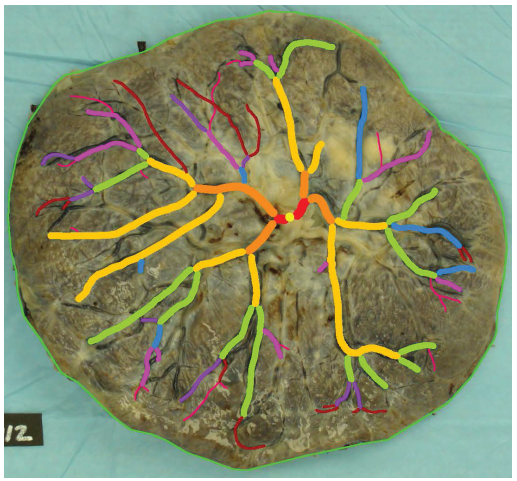
Traced

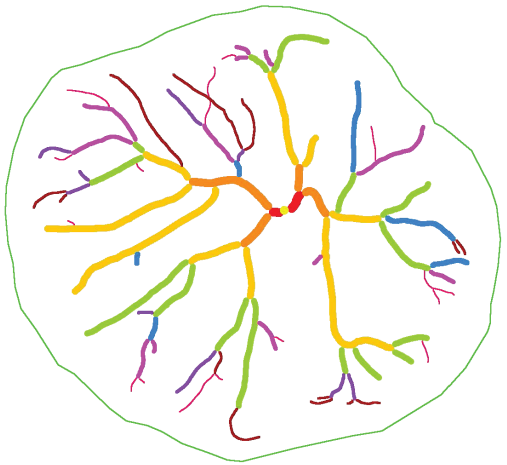
Arterial Network





Venous Network

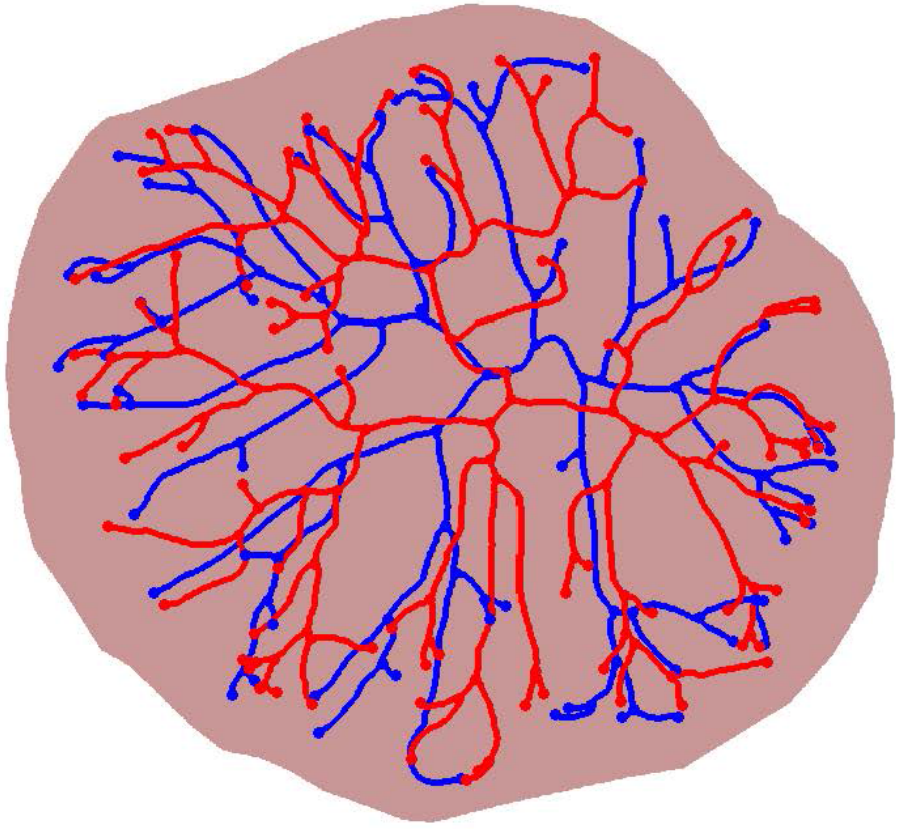




Skeletonized

```
%% Skeltrace
for i = 1:length(file_list)/2                                          % Till all files i
    filename = char(strcat(dirname,'\',file_list(i)));                  % Create filename
    [d, metric_area, metric_perimeter, perim, cimg] = skeltrace_test1(filename); % Obtaining s
    str =['Running ',num2str(i),' of ',num2str(length(file_list)/2),' files...'];
    disp(str)
    o = zeros(1,length(headers) - 1);

    %% Arteries
    [arteries_adjacency, arteries_skel_img, arteries_gpts, arteries_branchpts, arteries_endpts
     g, chains, tortuosity, v_area, perimeter] = new_skel_vesselprops(filename,d,perim);  % Co
    skeleton(1).trace_filename = char(file_list(i));                   % Name of arterial
    skeleton(1).type = 'arteries';                                     % Identify type of
    skeleton(1).chains = chains;                                       % Edge parameters
    skeleton(1).graph_points = g.graphpoints;                          % Graphpts. of art
    skeleton(1).angle = skelNearAngleCompute(skeleton(1));                 % Branching ar
    skeleton(1).tortuosity = tortuosity;                               % Arterial tortuos
    skeleton(1).VesselToDiscPercent = (single(v_area)/metric_area) *100;   % Percent of arter
    [~,fname] = fileparts(filename);                                   % Filename
    fname = fname(4:end);                                              % Removing arteria
    o(1) = v_area;                                                     % Arterial surface
    o(2) = skeleton(1).VesselToDiscPercent;                            % Percent of arter
    branches = length(skeleton(1).chains);                             % No. of arterial
    n_arteries = branches;                                             % No. of arteries
    gen = zeros(branches,1);                                           % Initialize a vec
    dia = zeros(branches,1);                                           % Initialize a vec
    len = zeros(branches,1);                                           % Initialize a vec
    vol = zeros(branches,1);                                           % Initialize a vec
    for i1 = 1:branches                                                % For each artery
        gen(i1) = skeleton(1).chains(i1).generation;                       % Save its ger
        dia(i1) = skeleton(1).chains(i1).diameter;                         % Save its dia
        len(i1) = skeleton(1).chains(i1).arc_length;                       % Save its arc
        vol(i1) = pi*(dia(i1)/2)^2 * len(i1);                                  % compute
    end
    o(3) = max(gen);                                                   % Find max arteria
    o(4) = nnz(gen == 1);                                              % No. of arterial
    o(6) = ceil((branches + o(4))/2);                                  % Compute no. of a
    o(5) = length(skeleton(1).graph_points) - o(6);                    % Compute no. of a
    o(7) = sum(len);                                                   % Arterial arcleng
    o(8) = sum(vol);                                                   % Arterial voulme
    o(9) = mean(dia);                                                  % Mean arterial di
```



PCSVN features

64 PCSVN features: 8 on shape, 56 on vessels (1/2 arterial & 1/2 venous)

BW_g	PW_g	GA_wk	Gender
3713.85	530.00	39.10	2.00
Calculated Beta	**Area**	**Perimeter**	**Compactness**
0.76	364.66	80.54	0.71
Eq_Circle_IP_Symmetry	**Sigma_UCI**	**Rmean**	**RmeanN**
0.33	2.03	10.49	0.55
UCI_to_Perim	A_SurfaceArea	A_VesselToDiscPercent	A_NumGenerations
7.52	32.52	8.92	12.00
A_NumCordBranches	A_NumBranchPoints	A_NumEndPoints	A_ArcLength
2.00	70.00	71.00	200.75
A_Volume	A_MeanThickness	A_StdThickness	A_MurrayBranchesUsed
5.73	0.16	0.10	69.00
A_MurrayExponent	A_MurrayL1FitError	A_MeanDistToPerim	A_StdDistToPerim
1.69	0.07	3.88	1.77
A_MeanDistEndPointToPerim	A_StdDistEndPointToPerim	A_MeanAngle	A_StdDevAngle
2.77	1.62	98.20	18.85
A_ModeAngle	A_MaxAngle	A_MinAngle	A_MedianAngle
90.00	143.00	56.00	95.50
A_KurtosisAngle	A_MeanTortuosity	A_StdDevTortuosity	A_MaxTortuosity
2.61	1.11	0.04	1.25
A_MinTortuosity	A_KurtosisTortuosity	V_SurfaceArea	V_VesselToDiscPercent
1.04	4.64	45.06	12.36
V_NumGenerations	V_NumCordBranches	V_NumBranchPoints	V_NumEndPoints
11.00	3.00	64.00	66.00
V_ArcLength	V_Volume	V_MeanThickness	V_StdThickness
215.41	9.59	0.21	0.11
V_MurrayBranchesUsed	V_MurrayExponent	V_MurrayL1FitError	V_MeanDistToPerim
63.00	1.88	0.08	3.70
V_StdDistToPerim	V_MeanDistEndPointToPerim	V_StdDistEndPointToPerim	V_MeanAngle
2.12	2.83	1.78	99.47
V_StdDevAngle	V_ModeAngle	V_MaxAngle	V_MinAngle
16.27	90.00	138.00	72.00
V_MedianAngle	V_KurtosisAngle	V_MeanTortuosity	V_StdDevTortuosity
101.00	2.58	1.11	0.07
V_MaxTortuosity	V_MinTortuosity	V_KurtosisTortuosity	
1.74	1.05	53.06	

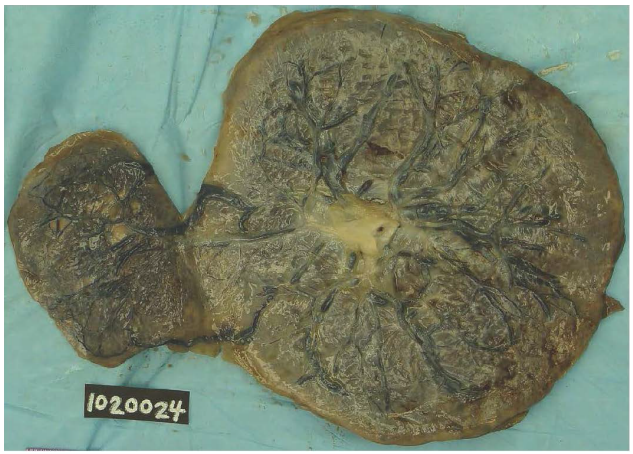
Standard deviation of radii wrt mean radius obtained from UCI

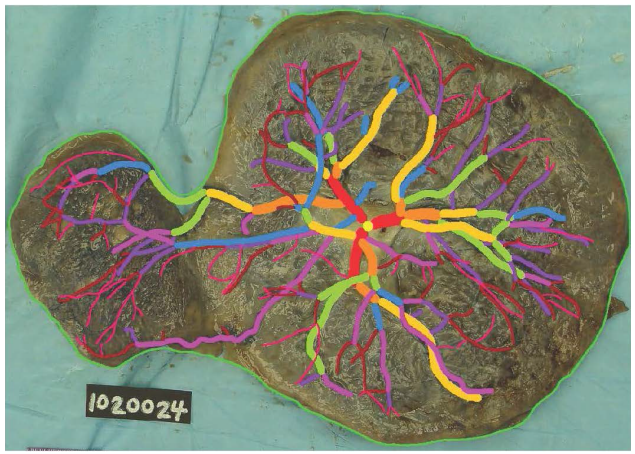
Total number of arterial end points

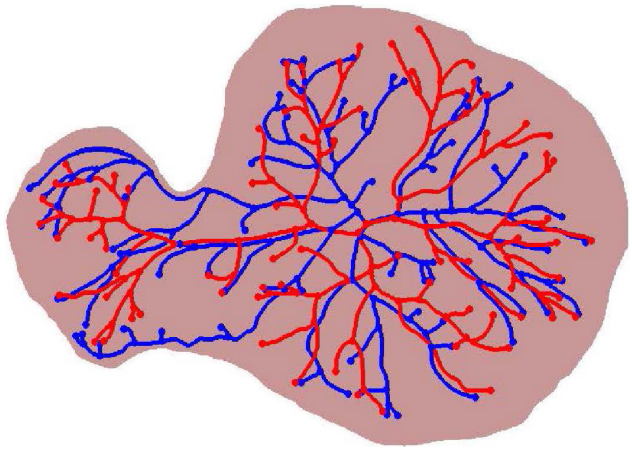
mean of all braching angles of venous network

Each placenta has a geometric signature

**Results are presented on the arterial network only since the arterial vessels were traced with a higher level of precision and accuracy* [4].

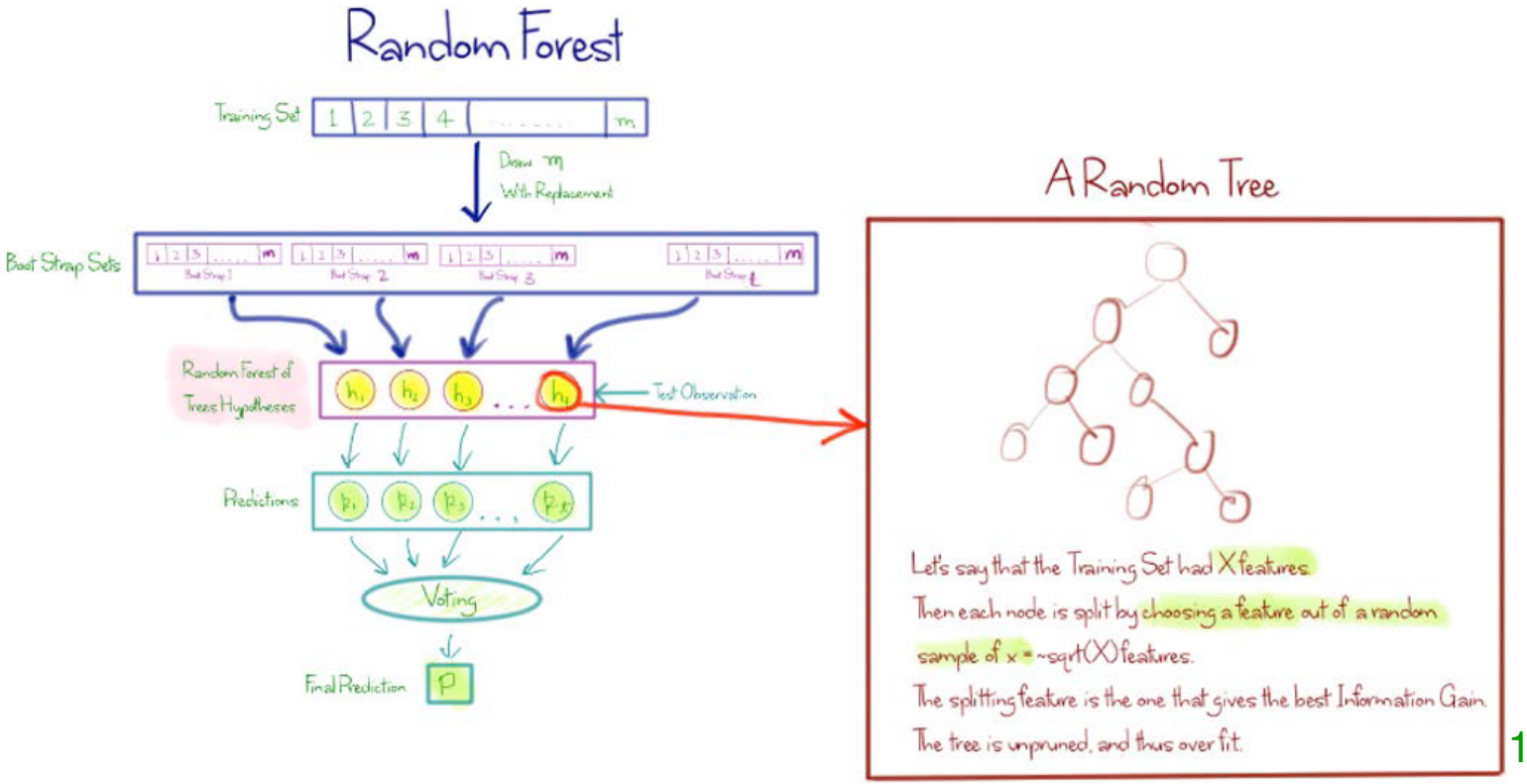






BW_g	PW_g	GA_wk	Gender	Calculated Beta
3714	530	39.1	2	0.763
Area	Perimeter	Compactness	Eq_Circle_IP_Symmetry	Sigma_UCI
364.658	80.543	0.706	0.332	2.029
Rmean	RmeanN	UCI_to_Perim	A_SurfaceArea	A_VesselToDiscPercent
10.486	0.549	7.524	32.52432127	8.919120789
A_NumGenerations	A_NumCordBranches	A_NumBranchPoints	A_NumEndPoints	A_ArcLength
12	2	70	71	200.75214
A_Volume	A_MeanThickness	A_StdThickness	A_MurrayBranchesUsed	A_MurrayExponent
5.733491374	0.161212422	0.09604111	69	1.689969339
A_MurrayL1FitError	A_MeanDistToPerim	A_StdDistToPerim	A_MeanDistEndPointToPerim	A_StdDistEndPointToPerim
0.065766908	3.882499448	1.765775945	2.772167898	1.620571594
A_MeanAngle	A_StdDevAngle	A_ModeAngle	A_MaxAngle	A_MinAngle
98.2	18.85305496	90	143	56
A_MedianAngle	A_KurtosisAngle	A_MeanTortuosity	A_StdDevTortuosity	A_MaxTortuosity
95.5	2.612210204	1.108739853	0.036899481	1.248528242
A_MinTortuosity	A_KurtosisTortuosity	V_SurfaceArea	V_VesselToDiscPercent	V_NumGenerations
1.036478162	4.637265682	45.05599548	12.35567284	11
V_NumCordBranches	V_NumBranchPoints	V_NumEndPoints	V_ArcLength	V_Volume
3	61	66	215.4093653	9.585888769
V_MeanThickness	V_StdThickness	V_MurrayBranchesUsed	V_MurrayExponent	V_MurrayL1FitError
0.212766845	0.11003167	63	1.879110501	0.075148675
V_MeanDistToPerim	V_StdDistToPerim	V_MeanDistEndPointToPerim	V_StdDistEndPointToPerim	V_MeanAngle
3.702251262	2.115948672	2.826950523	1.78410397	99.46960697
V_StdDevAngle	V_ModeAngle	V_MaxAngle	V_MinAngle	V_MedianAngle
16.27477491	90	138	72	101
V_KurtosisAngle	V_MeanTortuosity	V_StdDevTortuosity	V_MaxTortuosity	V_MinTortuosity
2.578977194	1.11495626	0.0686891	1.738315463	1.052159905
V_KurtosisTortuosity				
53.06433487				

Decision trees, random forest, and Boruta strategy



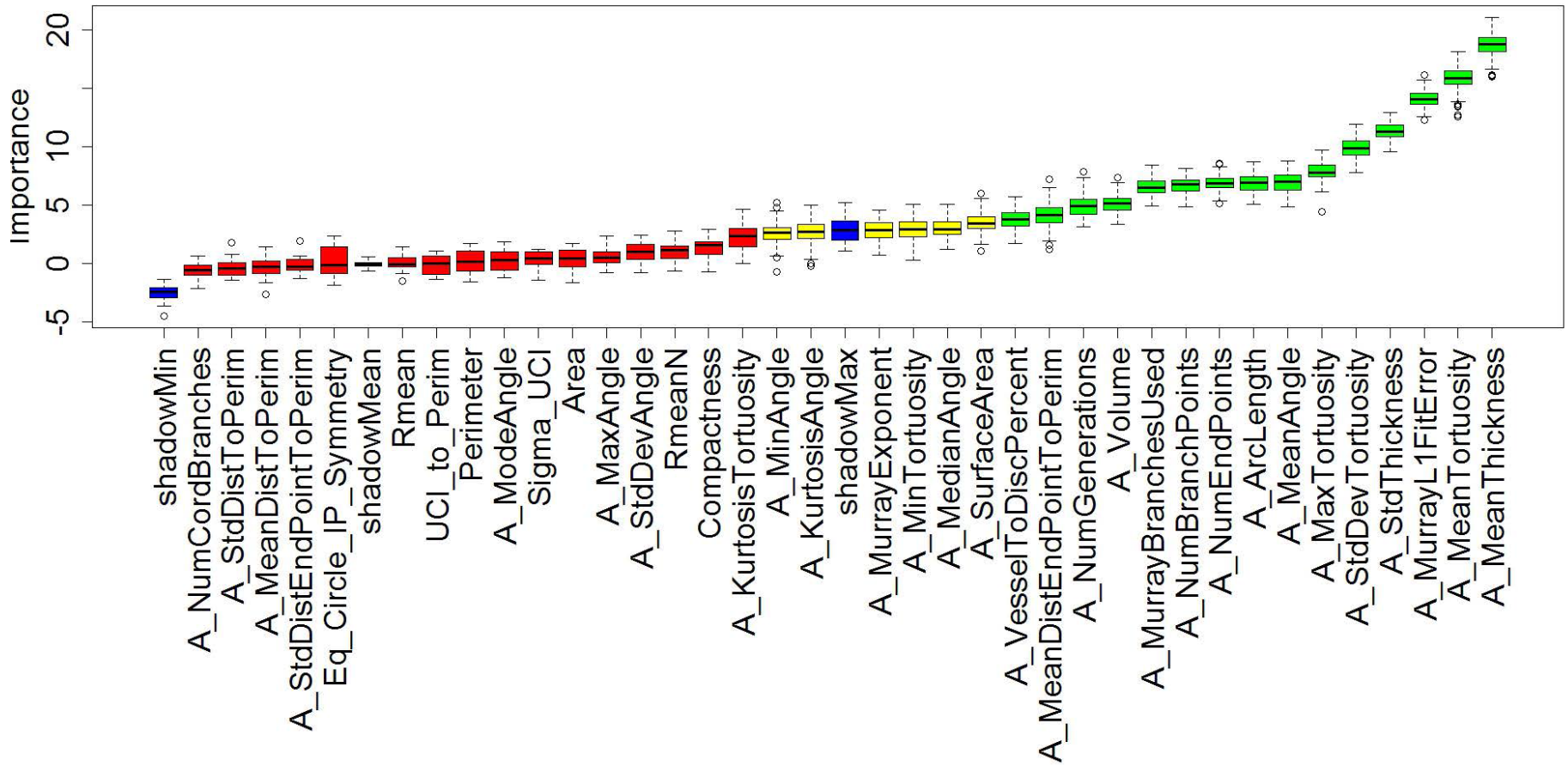
1

- In Boruta, each attribute has a *shadow attribute (SA)*, created by shuffling the values of the original attribute. • An attribute's importance is determined by comparing to the maximum z-score among all SA (MZSA).

[1] https://helloacm.com/a-short-introduction-bagging-and-random-forest/

Relevant feature selection with Boruta strategy

15 vessel-based features were selected after running random forest classification 500 times.



Principal component analysis (PCA) for best basis

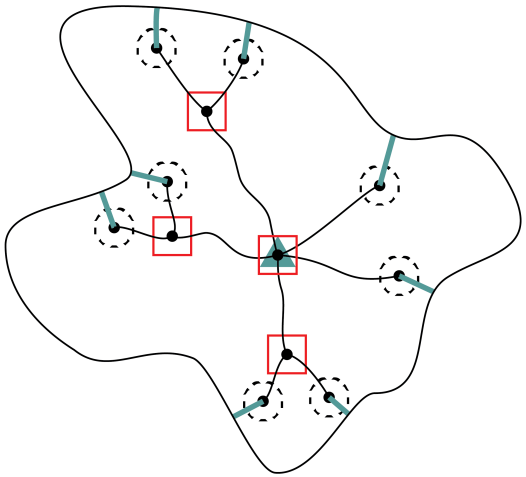
1. Let $F = \begin{bmatrix} | & | & & | \\ \mathbf{f}_1 & \mathbf{f}_2 & \cdots & \mathbf{f}_N \\ | & | & & | \end{bmatrix}$ be the $15 \times N$ data matrix, where $N =$ number of samples (290 in this case).
2. The 15×15 covariance matrix $C = \frac{1}{N-1}\tilde{F}\tilde{F}^T$ gives the feature variance (diagonal) and co-variance (off-diagonal), where $\tilde{F} = F - \mathbf{m}$ and $\mathbf{m}$ is the data mean.
3. Factor C through the singular value decomposition (SVD) to extract the basis of F in U (i.e., eigen-decomposition of C):
   $C = USV^T \quad (CU = SU \Rightarrow C = SUU^T \Rightarrow C = USU^T)$.
4. Project the 15-d data onto a 5-d ($\approx 88\%$ variance) space using the first 5 vectors in U:
   $$D = U(:, 1:5)^T \tilde{F}$$
   This is the space where classification takes place.

Dimensionality reduction with PCA

The first five principal components (PCs) of the correlation matrix retain approximately 88% of the data variance. The absolute value of the attributes within each PC gives a measure of contribution. The higher the value, the bigger the contribution.

Boruta ranking	Vascular features (variance captured)	PC1 (35.27%)	**PC2** (22.57%)	PC3 (17.20%)	PC4 (7.79%)	PC5 (5.80%)
1	**MeanThickness**	−0.1582	**−0.4747**	0.1035	0.0651	−0.0089
2	MeanTortuosity	0.0002	0.0575	**0.5347**	−0.0979	0.0013
3	MurrayL1FitError	−0.256	−0.3903	0.0438	0.0139	0.0397
4	**StdThickness**	−0.1566	**−0.4762**	0.0701	−0.0046	0.0196
5	StdDevTortuosity	0.0029	0.0812	**0.5912**	−0.0641	0.1449
6	MaxTortuosity	0.0948	0.0724	**0.5459**	−0.0264	0.1709
7	MeanAngle	−0.0611	0.0704	0.2028	0.2135	**−0.936**
8	NumEndPoints	**0.4251**	−0.0298	−0.0132	0.0153	−0.005
9	ArcLength	0.3773	−0.1259	−0.0035	−0.0163	0.0116
10	NumBranchPoints	**0.4254**	−0.0301	−0.0125	0.0146	−0.0038
11	MurrayBranchesUsed	**0.4254**	−0.0301	−0.0125	0.0146	−0.0038
12	**Volume**	0.1444	**−0.4823**	0.065	0.0502	−0.0368
13	NumGenerations	0.3182	−0.0237	0.014	0.2178	−0.0619
14	MeanDistEndPointToPerim	0.0055	−0.0323	0.0545	**0.905**	0.2124
15	VesselToDiscPercent	0.255	−0.3502	0.0031	−0.2561	−0.1457
		Branching	**Thickness**	Tortuosity	Growth	Branch Angle

Visualization of discriminating features

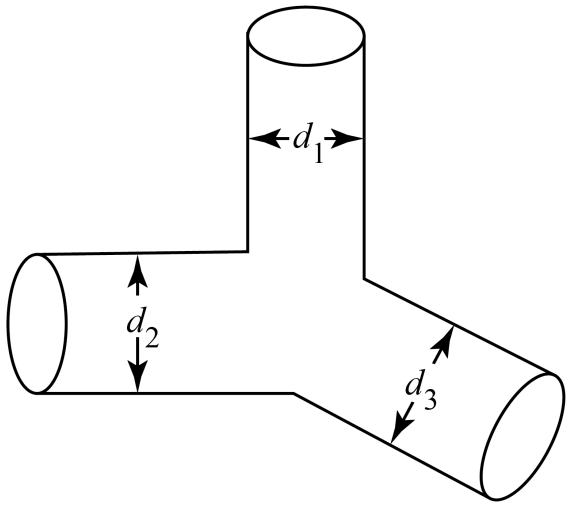


: Umbilical cord insertion (UCI)

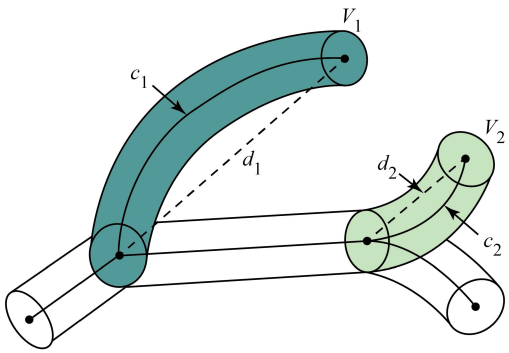
: End points

: Branch points

: Distance between an end point and its nearest boundary point



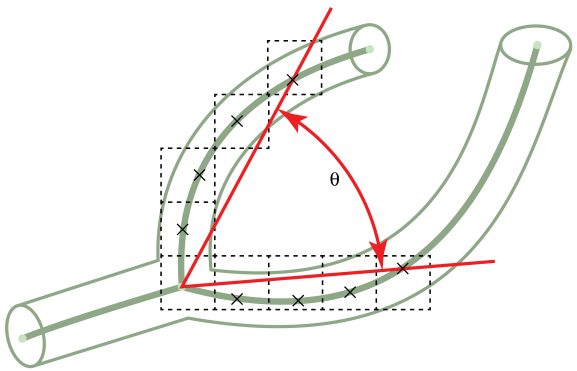
Thickness



c_i : Arc length of vessel i

d_i : Straight line distance between the initial and terminal nodes of vessel i

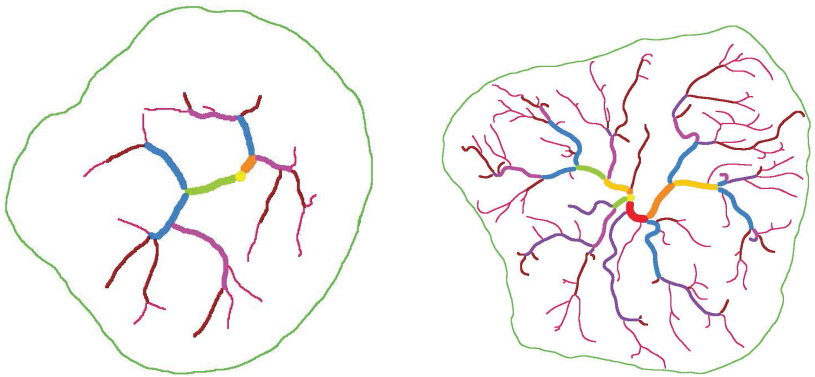
Tortuosity



Branching angle

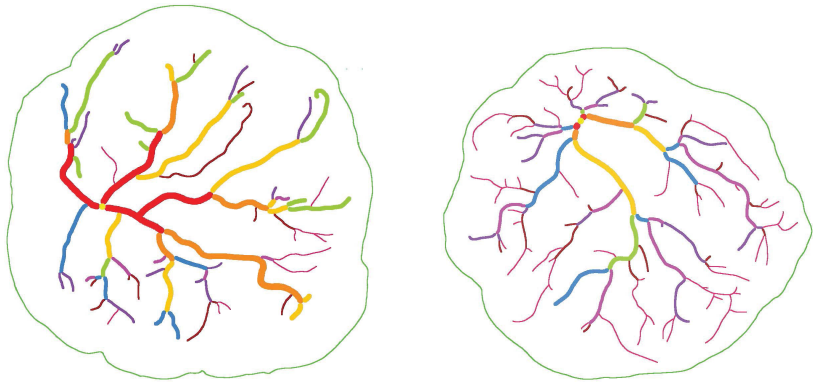
Visualization of PCSVN between 2 risk groups

high ASD risk low ASD risk



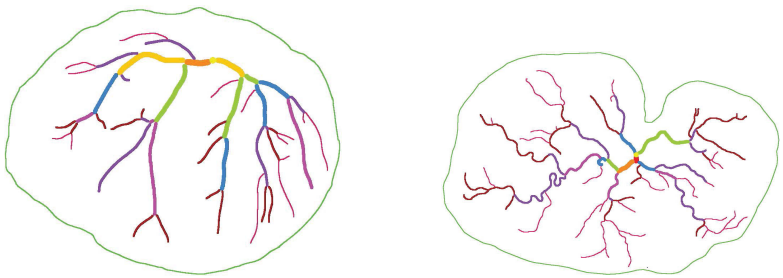
(a) Branching

high ASD risk low ASD risk



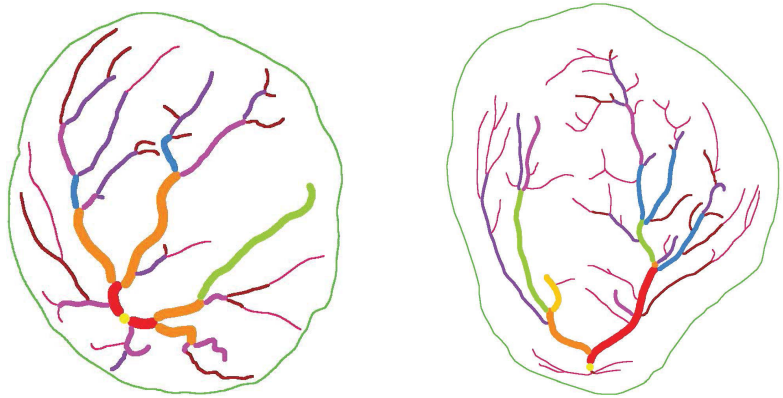
(b) Thickness

high ASD risk low ASD risk



(c) Tortuosity

high ASD risk low ASD risk



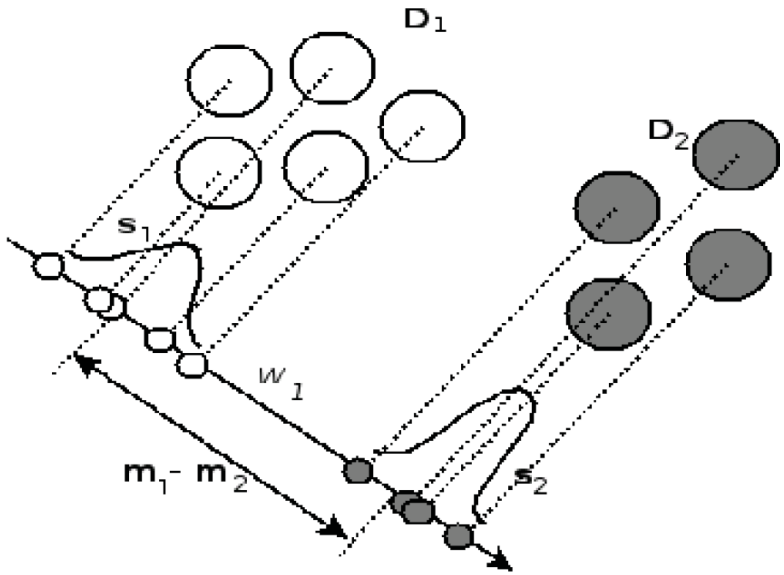
(d) Growth Extension

Create classification statistics with LDA

- Linear discriminant analysis amounts to finding a projection direction w_{opt} that maximizes the *between-class* scatter and minimizes the *within-class* scatter:

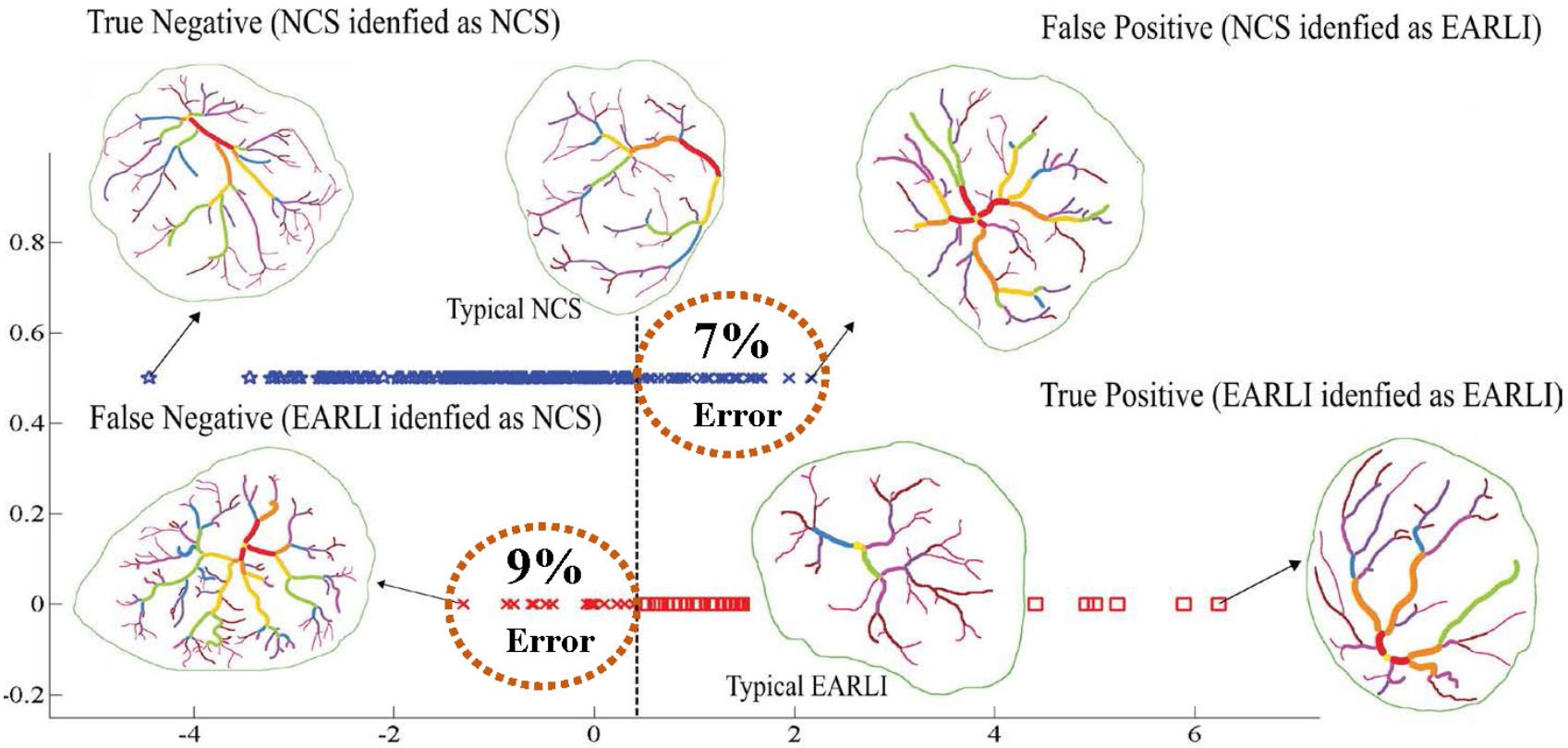
$$w_{\text{opt}} = argmax_{||w||=1} \frac{(\tilde{m}_2 - \tilde{m}_1)^2}{\tilde{S}_1^2 + \tilde{S}_2^2}$$

where $\tilde{S}_i^2 = \sum_{y \in D_i} (w^T y - \tilde{m}_i)^2$.



- In matrix form: $w_{\text{opt}} = argmax\, J(w)$, where $J(w) = \frac{w^T S_B w}{w^T S_W w} = \frac{N(w)}{D(w)}$.
- w_{opt} is the largest eigenvector associated with the largest generalized eigenvalue to the problem $N(w)w = \lambda D(w)w$.
- Set threshold $\alpha = \frac{1}{2}\left(\min\left\{w_{\text{opt}}^T D_2\right\} + \max\left\{w_{\text{opt}}^T D_1\right\}\right)$, then $C_1 = \left\{y \in D_1 \cup D_2 \mid w_{\text{opt}}^T y < \alpha\right\}$ and $C_2 = \left\{y \in D_1 \cup D_2 \mid w_{\text{opt}}^T y > \alpha\right\}$.
- Perform a 10-fold cross validation to generate classification statistics.

Visualization of classification statistics with LDA



Observations

- The difference in high and low ASD risk is better explained by the *vascular features alone*.
- PCSVNs associated with placentas of high-risk ASD pregnancies generally had *fewer branch points*, *thicker and less tortuous vessels*, *better extension to the surface boundary*, and *smaller branch angles* than their population-based counterparts.

Questions for which I have no answers ...

- What environmental or genetic factors cause this group of 5 parameters to vary together and whether these variables stabilize in their permanent state early in gestation?
- What types of geometric signatures that are measurable and capable of providing *accurate readings in 3-dimensional imaging environment*? (Answers to this question will play a vital role in early risk assessment and intervention for ASD.)
- The study presented here should motivate *a pursuit of additional PCSVN features*. What other shape signatures should be considered?
- *What PCSVN features are correlated* with ASD? We need *reliable and automated vessel extraction methods* to allow analysis of PCSVNs in large cohorts.

References

[1] G. Dawson, E. Jones, K. Merkle, K. Venema, R. Lowy, S. Faja, D. Kamara, M. Murias, J. Greenson, J. Winter, M.Smith, S. Rogers, S. Webb, "Early behavioral intervention is associated with normalized brain activity in young children with autism," *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(11), 1150–1159, 2012.

[2] J.-M. Chang, H. Zeng, R. Han, Y.-M. Chang, R. Shah, C. Salafia, C. Newschaffer, R. K. Miller, P. J. Katzman, J. Moye, M. Fallin, C. K. Walker, L. Croen, "Discriminating placentas of increased risk for autism with chorionic surface vascular network features," *under review*.

[3] C.J. Newschaffer, L.A. Croen, M.D. Fallin, I. Hertz-Picciotto, D.V. Nguyen, N.L. Lee, C.A. Berry, H. Farzadegan, H.N. Hess, R.J. Landa, S.E. Levy, M.L. Massolo, S.C. Meyerer, S.M. Mohammed, M.C. Oliver, S. Ozonoff, J. Pandey, A. Schroeder, K.M. Shedd-Wise, "Infant siblings and the investigation of autism risk factors," *J Neurodev Disord.*, 4(1), 1–16, 2012.

[4] R. Shah , C. Salafia, T. Girardi, L. Conrad, K. Keaty, A. Bartleotc, "Shape matching algorithm to validate the tracing protocol of placental chorionic surface vessel networks," *Placenta*, 36(8), 944–946, 2015.

[5] This talk is available via http://web.csulb.edu/~jchang9/files/PCSVN_forASDRik_JMC.pdf