Classification and Feature Selection for Craniosynostosis

Shulin Yang, Linda G. Shapiro, Michael L. Cunningham, Matthew Speltz, Su-In Lee

University of Washington
Seattle, WA, 98195, U.S.A.
{yang, shapiro, suinlee}@cs.washington.edu
{mcunning,mspeltz}@u.washington.edu

ABSTRACT

Craniosynostosis is the premature fusion of the bones of the calvaria resulting in abnormal skull shapes that can be associated with increased intracranial pressure. The goal of this work is to analyze the various 3D skull shapes that manifest in isolated single suture craniosynostosis. A logistic regression is used to identify different types of synostosis and quantify the differences. Due to the high-dimensionality of the feature data, a sophisticated feature selection technique is required to avoid overfitting and to improve the classification accuracy on the unseen data. In addition, feature selection allows the identification of surface areas that contribute to the major skull deformations that characterize isolated synostosis. We applied three sparse feature selection methods: L_1 regularization (lasso [9]), fused lasso ([10]) and a novel regularization method we have developed called the clustering lasso (cLasso). L_1 regularized logistic regression locates important surface points, and the fused lasso groups these points into regions. The cLasso was designed to assign similar weights to groups of correlated shape features. Experimental results indicated that the regularized logistic regression models achieve a significantly lower misclassification rate than unregularized logistic regression.

Categories and Subject Descriptors

J.3 [**Life and Medical Sciences**]; I.5.1 [**Pattern Recognition**]: Models; I.5.4 [**Pattern Recognition**]: Applications; I.4.9 [**Image Processing and Computer Vision**]: Applications

Keywords

craniosynostosis, cranial image (CI), L_1 penalized logistic regression, fused lasso, the clustering lasso (cLasso)

1. INTRODUCTION AND MOTIVATION

This paper is focused on 3D shape analysis of the skulls of patients with craniosynostosis. Craniosynostosis is a common congenital condition in which one or more of the fibrous sutures in an infant's calvaria fuse prematurely. This results in restricted skull and brain growth. Because the brain cannot expand perpendicular to the fused suture, it redirects growth in the direction of the open sutures, often resulting in an abnormal head shape and facial features. Some cases of craniosynostosis may result in increased intracranial pressure on the brain and developmental delays [8]. It is estimated that craniosynostosis affects 1 in 2,000 live births [7].

The motivation for this work is two-fold. First, as in previous work [3] [5], we propose methods for classification and quantification of three different types of isolated single suture craniosynostosis: coronal, metopic and sagittal. Second, in order to help physicians and researchers to better understand specific shape deformations, we pursue our analysis further to determine the local area of the skull in which the shape deformation was the most salient.

In our work, we built a system that automatically generates a shape representation called the cranial image (CI) [3] from the CT image of a patient's skull. The cranial images are used as features to distinguish between skulls of patients with different types of craniosynostosis. A logistic regression is used for this classification task, as well as for quantification of the shape deformation of the different types of craniosynostosis. Then, three variations of the logistic regression model are applied: L_1 regularized logistic regression [9], the fused lasso [10] and the clustering lasso (cLasso), which is proposed in this paper. These models select subsets of features from the cranial image, which represent skull regions that cause major shape differences among the three types of craniosynostosis.

2. RELATED LITERATURE

Calvarial (skull) abnormalities are frequently associated with severely impaired central nervous system functions due to brain abnormalities, increased intra-cranial pressure and abnormal build-up of cerebrospinal fluid. In [6], the authors introduced several different craniofacial descriptors that have been used in studies for two craniofacial disorders: 22q11.2 deletion syndrome (a genetic disorder) and deformational plagiocephaly/brachycephaly. They provided feature extraction tools for the study of craniofacial anatomy from 3D mesh data obtained from the 3dMD active stereo photogrammetry system. These tools produce quantitative representations (descriptors) of the 3D data that can be used to summarize the 3D shape as pertains to the condition being studied and the question being asked. This work is different from ours in that it analyzed the shape of the midface and

ACM-BCB '11, August 1-3, Chicago, IL, USA

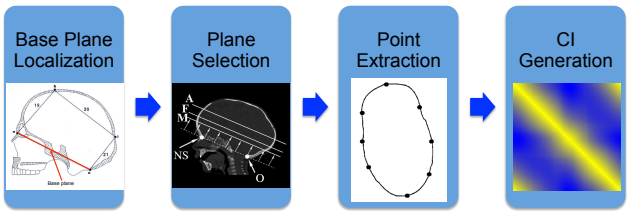

Figure 1: Framework of the system for automatic CI generation: with an input of 3D volume data in random pose, the system first extracts the skull and performs pose normalization, so that it is symmetric with respect to the right and left sides. Then, a base plane is located based on several landmarks that are biologically important. After that, one or multiple planes that are parallel to the base plane are extracted and a distance matrix (CI [3]) is composed from the pairwise distance between evenly spaced points on the skull contours.

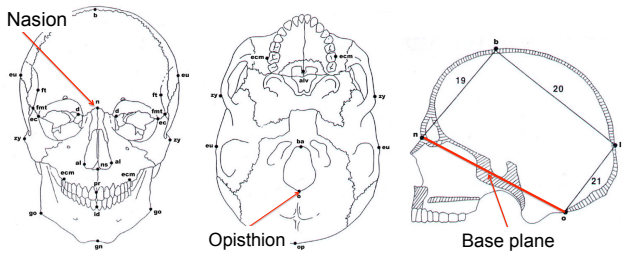

Figure 2: Nasion, opisthion and the base plane of the skull: the nasion is the intersection of the frontal and two nasal bones of the human skull [1]; the opisthion is the mid-point on the posterior margin of the foramen magnum on the occipital bone [1]; the base plane is the plane that goes through the nasion and the opisthion and is perpendicular to the middle plane of the head.

back of the head, while our work focuses on the shape of the skull.

Previous studies have also examined the specific skull shapes of patients. In [3] a novel approach to efficiently classify skull deformities caused by metopic and sagittal synostoses using symbolic shape descriptors was developed. In [5] a novel set of scaphocephaly severity indices (SSIs) for predicting and quantifying head- and skull-shape deformity in children diagnosed with isolated sagittal synostosis (ISS) was described, and their sensitivity and specificity was compared with those of the traditional cranial index.

These efforts differ from our approach in that they were not fully automatic and therefore required human interaction for selecting planes and landmarks from the skull for the purpose of extracting shape features. They use machine learning tools, but not a logistic regression approach.

3. CRANIAL IMAGE (CI) GENERATION

3.1 System Framework

A system was built for automatic generation of the cranial image [3]. The framework of the system is shown in Fig. 1. The following sections give a detailed explanation of each module of the system.

3.2 Base Plane Localization

In the first module, a base plane is located on the skull based on two important landmarks: the nasion and the opisthion (Fig. 2). The nasion is the intersection of the frontal and two nasal bones of the human skull [1]. Its manifestation on the visible surface of the face is a distinctly depressed area directly between the eyes, just superior to the bridge of the nose. The opisthion is the mid-point of the posterior margin of the foramen magnum on the occipital bone [1]. The two points were chosen because of their significance for analysis of the human skull, and because they are stable during the human growth process.

The approach to finding the nasion and opisthion includes several steps. First, the plane of symmetry of the left and right sides of the skull is extracted. The landmarks are expected to be on the intersection of this plane with the skull, which forms an outline as shown in Fig. 1. Then, three points are found; the tip of the nose, the nasion and the opisthion. These three points are detected in an order such that the location of each point is based on that of the previous one. The tip of the nose is first and is located by finding the point with the smallest horizontal value (x value in the Cartesian coordinate system) of the skull outline. The nasion is located as the point closest to the tip of the nose, which is above it and which has a zero curvature in the vertical direction. The opisthion is located as the point in the left part of the outline, which has the closest distance to the nasion of all points below the nasion. After that, a base plane is identified as the one that goes through the nasion and the opisthion and is perpendicular to the plane of symmetry of the head, as shown in Fig. 2.

3.3 Plane Selection

Our shape measure is based on the distances between points on the surface of the skull. The second module selects planes on which the points are located. The top plane of the skull is a plane that has intersection with the skull and which is parallel to the base plane but has the furthest distance to the base plane. Our system can extract any plane that is parallel to the base plane and located between the base plane and the top plane of the skull, based on the ratio of its distance to these two planes. Multiple planes may be selected and used together.

The selection of planes is related to our previous studies, which have proven the effectiveness of three planes in the skull: A-plane, F-plane and M-plane. The A-plane is at the top of the lateral ventricle, the F-plane is at the foramen of Munro, and the M-plane is at the level of the maximal dimension of the fourth ventricle [3]. In [3], these planes were manually selected using anatomical knowledge. In our approach, these three planes are automatically located between the base plane and the top plane of the skull with certain ratio values. The ratio values are estimated as 0.27, 0.36 and 0.54.

Our software allows users to select any planes according to their distance from the base plane. The three planes from [3] (A-plane, F-plane and M-plane) were used individually in the first set of experiments (section 5.2) in order to simulate

the experiments in [3]. Then in the rest of our experiments, 10 planes that are evenly distributed across the whole skull were used to provide a more general 3D shape descriptor.

3.4 Point Extraction and CI Generation

In the third and last module of the system, the outlines of the skulls are extracted on the planes from the previous module, and N points are evenly extracted along the outlines ($N = 100$). Pairwise distances among these points are calculated, forming an N by N distance matrix. The number at position (i, j) of the matrix represents the distance between point number i and point number j. The matrix is symmetric, thus the name cranial image (CI) [3].

4. SHAPE ANALYSIS USING CI

Our work on classification and shape quantification strongly differs from [3]. This section describes our methodology.

4.1 Logistic Regression for Classification

The cranial image is used for further analysis of the shape of the skulls. The first task is to classify the skulls into different types of craniosynostosis: coronal, metopic, and sagittal. Specifically, we consider all pixel values in a cranial image as features of a skull, and use logistic regression for classification.

Logistic regression is a workhorse in machine learning that uses a generalized linear model for binomial regression.

$$p(y|\mathbf{x}, \mathbf{w}) = \frac{1}{1 + exp(-y(\mathbf{w}^T\mathbf{x} + w_0))} \quad (1)$$

where vector $\mathbf{x}$ contains the feature values of a data sample; y is its class label (for example, $y = 1$ refers to coronal and $y = -1$ refers to metopic), $\mathbf{w}$ contains the coefficients for $\mathbf{x}$, and w_0 is the intercept. Furthermore, w_0 and $\mathbf{w}$ are model parameters, and $p(y|\mathbf{x}, \mathbf{w})$ is the probability that a data sample belongs to a certain class.

Coefficients w_0 and $\mathbf{w}$ in this model are the optimal parameters that minimize the following loss function:

$$l(w_0, \mathbf{w}) = \sum_{i=1}^{n} \log(1 + \exp(-y_i(\mathbf{w}^T\mathbf{x_i} + w_0))) \quad (2)$$

$$\{w_0, \mathbf{w}\} = \min_{w_0, \mathbf{w}} l(w_0, \mathbf{w}) \quad (3)$$

where y_i is the actual class label of a sample $\mathbf{x_i}$. The prediction of a sample data $\mathbf{x}$ is class label 1 if this criteria is true: $p(y = 1|\mathbf{x}, \mathbf{w}) > 0.5$.

4.2 L1 Regularized Logistic Regression

Due to the high-dimensionality of the data (i.e. a large number of features and a modest size of samples), learning the unregularized logistic regression [3] will result in overfitting. To avoid overfitting, we applied L_1 regularization that induces sparsity in the solution $\mathbf{w}$ such that many of the coefficients in $\mathbf{w}$ are set to exactly zero. L_1 regularization [9] has been rigorously proven to be effective in selecting relevant features when there are exponentially many irrelevant ones [2]. The log-likelihood of L_1 regularized logistic regression is as follows.

$$l(w_0, \mathbf{w}) = \sum_{i=1}^{n} \log(1+\exp(-y_i(\mathbf{w}^T\mathbf{x_i}+w_0)))+\lambda\sum_{i=1}^{m}|w_i| \quad (4)$$

where λ is a regularization parameter for the L_1-norm. Geometrically, each feature corresponds to two points on the skull surface; thus, the selected features - those corresponding to non-zero coefficients - can represent the skull points that contribute most to the shape deformation of the skulls.

4.3 The Fused Lasso for Selecting Feature Combinations

One problem with L_1 regularization for our purpose is that when features are highly correlated, it arbitrarily chooses one of many correlated features. Thus, although there are likely to be "regions" - involving multiple features that are physically close - that are predictive of the abnormalities, our results show that surface points corresponding to the selected features are usually scattered on the skull. This makes it hard to derive a medically relevant conclusion. Here, we propose to use a variation of L_1 regularization that better reflects the underlying structure of our feature data. Specifically, we use the fused lasso to induce bias such that the selected points that are close to each other form salient regions on the skull.

The fused lasso [10] places a constraint on the weights of the features that are geographically related - sharing the same or neighboring surface points. The loss function of the fused lasso is,

$$l(w_0, \mathbf{w}) = \sum_{i=1}^{n} \log(1 + \exp(-y_i(\mathbf{w}^T\mathbf{x_i} + w_0))) + \lambda\sum_{i=1}^{m}|w_i| + \mu \sum_{\{w_i, w_j\} \in M} |w_i - w_j| \quad (5)$$

where μ is a regularization parameter for the new penalty term. M is a set that contains all pairs of features that are neighbors, whose endpoints are the same or next to each other. In equation [5], $\lambda\sum_{i=1}^{m}|w_i|$ penalizes large feature weights, and $\mu\sum_{\{w_i,w_j\}\in M}|w_i - w_j|$ penalizes large weight differences between neighboring features. This new penalty term induces geographically related features to be assigned similar weights.

4.4 cLasso for Forming Feature Clusters

L_1 regularized logistic regression tends to assign different weights to highly correlated features. When features are highly correlated, it arbitrarily chooses one of them and assigns a non-zero weight only to it. The fused lasso places constraints on the weight differences base on their geographical relationships. Another option is to penalize the weight differences of correlated features. We propose a new form of regularized logistic regression, namely the clustering lasso (cLasso). The model for the clustering lasso is:

$$p(y|\mathbf{x}, \mathbf{w}, \mathbf{w^c}) = \frac{1}{1 + exp(-y(\mathbf{w}^T\mathbf{x} + \mathbf{w^c}^T + w_0))} \quad (6)$$

where $\mathbf{x}$ contains the feature values of a data sample; y is its class label (for example, $y = 1$ refers to sagittal and $y = -1$ refers to non-sagittal); $\mathbf{w}$ contains the coefficients for $\mathbf{x}$; $\mathbf{w^c}$ contains the coefficients for $\mathbf{c}$ (the cluster centers of $\mathbf{x}$); and w_0 is the intercept. Furthermore, w_0, $\mathbf{w}$ and $\mathbf{w^c}$ are model parameters, while $p(y|\mathbf{x}, \mathbf{w}, \mathbf{w^c})$ is the probability that a data sample belongs to a certain class.

Then the loss function for the cLasso becomes

$$l(w_0, \mathbf{w}, \mathbf{w^c}) = \sum_{i=1}^{n} \log(1 + \exp(-y_i(\mathbf{w}^T\mathbf{x_i} + \mathbf{w^c}^T\mathbf{c_i} + w_0))) + \lambda \sum_{i=1}^{m} |w_i| + \nu \sum_{i=1}^{k} |w_i^c| \quad (7)$$

where $\mathbf{c_i}$ ($i \in [1, k]$) is the centroid of a group of features $\{x_{i_1}, x_{i_2}, ..., x_{i_k}\}$ (its feature value is their average); w_i^c is the weight for c_i; and ν is the regularization parameter for the weights of the cluster centers.

This loss function is designed to cluster the features based on their correlation, and penalize their shared weights (w_i^c) and individual weights ($w_{i_1}, w_{i_2}, ..., w_{i_k}$) respectively. When ν is small and λ is large, individual weights are penalized, and features tend to be split into groups based on their correlation and to share the same weights. When λ is large enough, this model is equivalent to the model of L_1 regularized logistic regression (equation [4]).

Parameter w_i^c encourages correlated features to share the same weight, and w_i allows unique features to be used. Therefore, the cLasso is equivalent to using only shared weights (w_i^c) when each centroid c_i ($i \in [1, k]$) is computed as a weighted average with the weights determined by ($w_{i_1}, w_{i_2}, \ldots, w_{i_k}$).

5. EXPERIMENTS

5.1 Medical data

Our approach for analysis of skull shape was tested on 3D CT images of children's heads from hospitals in four different cities in the US. The children each had one of the three types of craniosynostosis: coronal, metopic or sagittal. In total we examined 70 CT images, each comprising a stack of image slices.

The experiments were designed to test the ability of our system to identify different types of craniosynostosis. Cranial images were generated using our system with one or multiple planes. Logistic regression, L_1 regularized logistic regression, the fused lasso and the cLasso were used for classification, quantification and interest region localization. In the following, we first tested the ability to classify different types of craniosynostosis using only one plane. Then multiple planes were used for classification and interest region localization. Regularization parameters in the models were analyzed, and classification results were compared using different models. Implementation of L_1 regularized logistic regression is from the authors of [2], and implementation of the fused lasso is the machine learning package SLEP [4].

5.2 Single-Plane Classification

In the first experiment, we tested the cranial images generated using only one plane: the A, F or M plane (introduced in section 3.3). 100 points were extracted on each plane to calculate the distance matrix. Misclassification rates were measured using 3-fold cross validation. A logistic regression model was used as the classifier.

The results of single-plane classification are shown in Table 1. The classifier was first tested on each pair of classes and then for all three classes at once, using all features. Based on the experimental results, the lowest misclassification rates are achieved using the A-plane or F-plane alone.

One Plane	**C** vs **M**	**M** vs **S**	**S** vs **C**	3-Classes
A-Plane	**3.29**%	**12.67**%	26.29%	10%
F-Plane	4.39%	17.57%	**25.57**%	10%
M-Plane	6.29%	17.14%	27.14%	10%

Table 1: Misclassification rates using a single plane: C represents coronal; M represents metopic; S represents sagittal. Classification is run for both pairs of classes and multiple classes.

Although the three planes are biologically meaningful, the misclassification rates of sagittal versus coronal are high, over 25% using any of the 3 planes. Improvements are needed for the results to be able to help with craniosynostosis research.

5.3 Parameter Selection

In the second experiment for multi-plane classification, the cranial images were generated from 10 planes that were evenly spaced across the skull. 10 points were extracted from each plane, and a cranial image with 10,000 features was used for classification using logistic regression and its variations with different regularization terms. Misclassification rates were measured using 3-fold cross validation.

The first problem in conducting these tests was to determine the value of the regularization parameters λ, μ and ν in equations [4], [5] and [7]. The effect of these regularization parameters on classification accuracy was explored by testing the misclassification rate of sagittal versus coronal with fixed training and testing sets. Fig. 3 examines the relationship of λ, μ, ν and the misclassification rate in equation [4], [5] and [7]. Based on the observation in the figures, the classification error is the lowest when $0.5 \leq \lambda \leq 1$, $\mu \approx 0.2$ and $\nu < 0.1$.

In the classification task, the regularization parameters were found using 10-fold cross validation on the training set. Specifically, the training set for each test was divided into 10 sub-folds. Nine of them were used for training with certain regularization parameters, and the other fold was used for testing the misclassification rate. The parameter setting with the lowest average rate across all 10 sub-folds was chosen.

5.4 Classification Results & Visualization

Classification results using equally-spaced boundary points from 10 planes with the four different models are shown in Table 2. The results of logistic regression are modest despite the large number of features it uses. This substantiates the overfitting problem when using this model. The misclassification rate greatly improves when regularization is used in the logistic regression model. Specifically, the clustering lasso, which was proposed in this paper, exhibits a significant improvement in the misclassification rate, particularly for the sagittal class. The interest region localization results using the fused lasso are shown in Fig. 4.

6. CONCLUSIONS

In this work, we built a system that performed automatic analysis of skull shape. A shape measure called the cranial image (CI) is automatically generated by our system. Logistic regression, L_1 regularized logistic regression, the fused lasso and the clustering lasso, which is a model proposed in

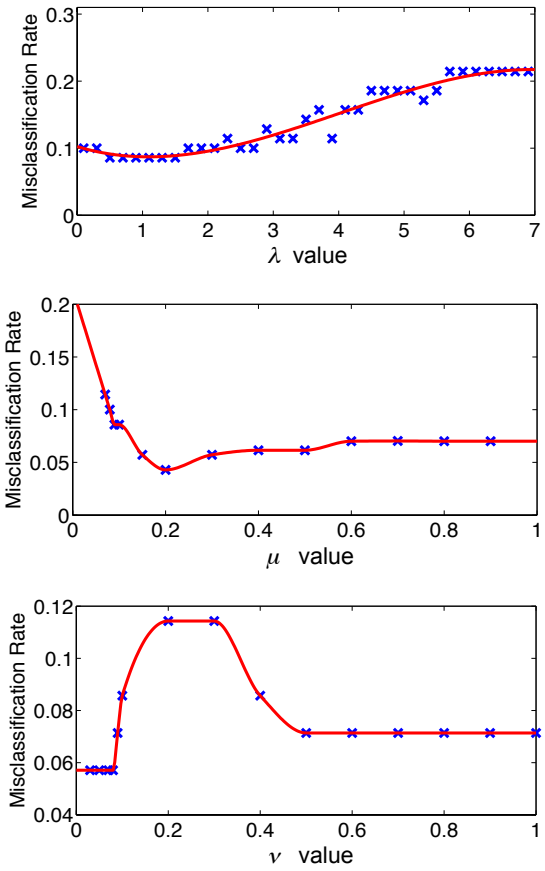

Figure 3: Misclassification rate vs regularization parameters: the x axis is value of regularization parameter λ, μ and ν; the y axis is the misclassification rate achieved using the value.

this paper, are used for classification of different types of isolated single suture craniosynostosis. Our experimental results show a significant improvement in the misclassification rate using the model we proposed. The clustering lasso is also used for quantification of different syndromes. L_1 regularized logistic regression and the fused lasso are used to select features from the CI. With these selected features, the most representative deformed shape regions can be shown on the surfaces of patient skulls.

We expect our methodology to be useful in craniofacial research. The interest region location and quantification results help researchers interested in clinical outcomes research. The classification results may provide a convenient tool to assist with remote diagnostics.

7. ACKNOWLEDGMENTS

This research was supported by NIH/NIDCR under grant number U01 DE 020050 (Dr. Shapiro) and grant number R01 DE 13813 (Dr. Speltz), and by the Jean Renny endowment for Craniofacial Medicine (Dr. Cunningham).

Multiple Planes	C vs M	M vs S	S vs C	3-Classes
Logistic regression	13.57%	13.57%	23.93%	10%
L_1 regression	7.14%	5%	6.43%	8.57%
Fused lasso	5.71%	5.71%	**4.29%**	18.57%
Clustering lasso	**4.29%**	**4.29%**	5.71%	**7.14%**

Table 2: Misclassification rates using multiple planes: C represents coronal; M represents metopic; S represents sagittal. Classification is run for both pair of classes and multiple classes.

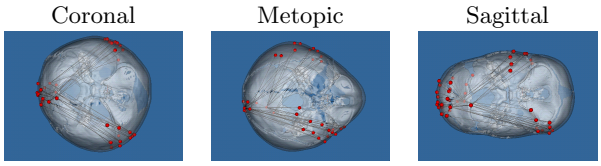

Figure 4: Visualization of selected points together with their pairwise distances using the fused lasso.

8. REFERENCES

[1] H. Gray and H. V. Carter. *Gray's Anatomy*. Sterling Publishing, 2000.

[2] S.-I. Lee, H. Lee, P. Abbeel, and A. Y. Ng. Efficient L1 regularized logistic regression. *Proceedings of the 21st National Conference on Artificial Intelligence*, 2006.

[3] H. Lin, S. Ruiz-Correa, R. Sze, M. Cunningham, M. Speltz, A. Hing, and L. Shapiro. Efficient symbolic signatures for classifying craniosynostosis skull deformities. *Workshop of ICCV*, 2005.

[4] J. Liu, S. Ji, and J. Ye. *SLEP: Sparse Learning with Efficient Projections*. Arizona State University, 2009.

[5] S. Ruiz-Correa, R. Sze, J. Starr, H. Lin, M. Speltz, M. Cunningham, and A. Hing. New scaphocephaly severity indices of sagittal craniosynostosis: A comparative study with cranial index quantifications. *Cleft Palate-Craniofacial Journal*, 2006.

[6] L. Shapiro, K. Wilamowska, I. Atmosukarto, J. Wu, C. Heike, M. Speltz, and M. Cunningham. Shape-based classification of 3d head data. *ICIAP*, 2009.

[7] B. J. Slater, K. A. Lenton, M. D. Kwan, D. M. Gupta, D. C. Wan, and M. T. Longaker. Cranial sutures: a brief review. *Plastic and Reconstructive Surgery*, 121(4):170–178, 2008.

[8] J. R. Starr, K. Kapp-Simon, Y. K. Cloonan, B. R. Collett, M. M. Cradock, L. Buono, M. L. Cunningham, and M. L. Speltz. Pre- and post-surgery neurodevelopment of infants with single-suture craniosynostosis: Comparison with controls. *Journal of Neurosurgery (Pediatrics)*, 107(2):103–110, 2007.

[9] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 58(1):267–288, 1996.

[10] R. Tibshirani, M. Saunders, S. Rosset, Y. Heights, J. Zhu, and K. Knight. Sparsity and smoothness via the fused lasso. *J. R. Statist. Soc. B*, 67:91–108, 2005.