

Insight Into Lumbar Back Pain

What the Lumbar Spine Tells About Your Life

Paul Klemm¹, Sylvia Glaßer¹, Kai Lawonn¹, Marko Rak¹ Henry Völzke², Katrin Hegenscheid³ and Bernhard Preim¹

¹Department of Simulation and Graphics, University of Magdeburg

²Institut for Community Medicine, University of Greifswald

³Institute for Diagnostic Radiology and Neuro-Radiology, University of Greifswald

{paul, sylvia}@isg.cs.uni-magdeburg.de, {katrin.hegenscheid, voelzke}@uni-greifswald.de

Keywords: Epidemiology, Interactive Visual Analysis

Abstract: The goal of epidemiology is to provide insight into disease causations. Hence, subject groups (cohorts) are analyzed to correlate the subjects' varying lifestyles and medical properties with diseases. Recently, these cohort studies comprise medical image data as well. We assess the predictive quality of image-derived variables for lower back pain. The evaluation is embedded in an interactive visual analysis (IVA) framework that is based on 2,540 segmented lumbar spine data sets. The segmentation results are evaluated and quantified by employing shape-describing variables, such as the spine canal course, its curvature and torsion. We analyze the predictive power of shape-describing variables for non-image variables, e.g., pain indicators. Therefore, we automatically extract a decision tree classifier for each non-image variable. Furthermore, we provide an IVA technique to compare the classifiers with interactive scatterplot visualizations. As a first result, we conclude that image-based variables are only sufficient for classification of lifestyle factors. A correlation between lumbar spine shape and lower back pain disease was not existent in the automatically trained classifiers. However, the presented approach is a valuable extension for the IVA of epidemiological data. Hence, correlations between non-image parameters were successfully detected and characterized.

1 INTRODUCTION

Epidemiology is the study of dissemination, causes and results of health-related states and events. For this purpose, large population studies, such as the *Study of Health in Pomerania* (SHIP) (Völzke et al., 2011) gather as much information as possible about participants to be assessed towards different diseases. These information are used to determine risk factors for diseases, help people to improve their lifestyle and their health or to support diagnosis of widespread diseases. Epidemiological research is strongly hypothesis-driven. More precisely, observations made by clinicians are translated into hypotheses, which are then statistically evaluated using data variables from epidemiological studies. As a result, the hypothesis will be accepted or rejected.

For epidemiological research, cohort studies are carried out. Since recently, these studies often comprise medical image data, e.g., tomographic image data sets, as well as general patient-specific attributes, such as *gender*, *age* and *weight*. Back pain is one of

the most frequent diseases in the western civilization and gains importance due to the aging society. Epidemiological analysis of back pain aims to characterize the disease by determining risk factors. Although the shape and constitution of the spine, especially the lumbar spine, plays an important role for the presence of back pain, no automatic classification approach to characterize back pain based on lumbar back spine attributes has been established. Furthermore, previous analyses mainly focused on a single parameter or a small subset of parameters.

In this paper, we present a thorough analysis of the image-derived data from lumbar back spine datasets and analyze their discriminating power for different characteristics of the disease. For this purpose, we combine data mining algorithms with data visualization techniques. Then, interactive visual analysis provides insight into the quality of the image-derived data as well as their potential for being back pain risk factors. The workflow is part of an interactive visual analysis (IVA) framework, constructed to enhance the epidemiological workflow with iterative data driven

techniques to derive insight (Klemm et al., 2014). Our contributions are:

- An IVA workflow for back pain analysis based on image-derived variables of 2,240 subjects
- The identification and evaluation of lumbar spine shape properties for back pain diagnosis
- Detection of correlations between image-based and socio-demographic as well as medical parameters for an improved hypothesis generation
- Identification of the most important variables via data mining methods (including a novel semi-quantitative evaluation concept) embedded in our IVA-framework

2 EPIDEMIOLOGICAL BACKGROUND

Epidemiological reasoning relies on a strict statistically driven workflow (Fletcher et al., 2012):

- Physicians formulate hypotheses based on observations made in their clinical practice.
- To assess a hypothesis, epidemiologists compile a list of variables depicting it.
- Statistical methods, such as regression analysis, assess the association of selected variables with the investigated disease.

Mutually dependent variables make this analysis challenging. Many diseases, such as different cancer types, are more likely with increasing age. When for example analyzing influences of nutrition to prostate cancer, the results need to be age-normalized. Age acts as a *confounder* for prostate cancer. Statistical correlation does not imply causation—epidemiologists need to assess the medical soundness of the statistical results.

2.1 Epidemiological Data

Epidemiological data originates from a wide range of studies. The study type depends on the condition of interest. Most common are case-control studies, analyzing one specific disease and its influences on the human body. We focus on data from large scale cohort studies. These studies aim to collect as much data as possible for each subject. As a result, these data can be analyzed regarding many diseases and conditions.

Epidemiological data is heterogenous and incomplete. For example women-specific questions or data about a disease treatment only subjects suffering from

this condition. Statistical analysis has to take missing data into account.

Epidemiologists acquire data using a wide range of techniques, such as medical examinations, self reported questionnaires or genetic examinations. This yields a heterogenous information space. To compare these data, information reduction techniques are applied. For example, continuous data, such as age is often discretized into age-bins. Every information reduction can introduce a bias to the data, since it reflects an assumption about the data. Using age to divide subjects into *young* and *old* categories can distort statistical results. This distortion is reduced with increasing number of discretization steps.

Modern cohort studies often comprise medical image data. These data are hard to analyze. Individual diagnosis or manual segmentation of each body structure by an radiologists is tedious, costly and comprises little reproducibility. Segmentation algorithms are not generally available and need to be custom-made for each body structure. Segmentation data is usually analyzed by abstracting it into key figures, such as diameters or distances. These numeric values can be compared with non-image variables to retrieve correlations.

2.2 Lower Back Pain

The lower (*lumbar*) spine is the most stressed spine part. Lower back pain is one of the most frequent diseases in the western civilization. Epidemiologists assume associations between lumbar back pain and lifestyle factors. These include nutrition, sporting activities and body posture combined with physical stress at work. The exact causes as well as particularly vulnerable risk groups are not known. Potential *confounding* effects are also subject of current research.

Epidemiologists want to characterize the healthy aging process of the spine. To achieve this, they have to analyze the lumbar spine shape as well as the mentioned lifestyle factors.

3 RELATED WORK

Visual Analysis of Image and non-Image Data
Klemm et al. propose an extension to the previously described epidemiological workflow using *Interactive Visual Analysis* (IVA) (Klemm et al., 2014). The workflow consists of an iterative sequence of *group selection*, *variable selection* and *visualization*. It aims to trigger *hypothesis generation* by providing visualizations able to concurrently analyze multiple heterogenous variables at once using correlation mea-

sures. We incorporate the IVA approach by concurrently displaying complex relationships between variables using decision trees. Steenwijk et al. propose a coordinated linked view system for both image-related and non-image data (Steenwijk et al., 2010). It incorporates multiple plot types, such as scatter plots, parallel coordinates and time plots with brushing and linking facilities. Similar to our work, they quantify image data and project it into the non-image data information space. Turkay et al. follow a similar approach by deriving descriptive data metrics from image data (Turkay et al., 2013). Their proposed *deviation plot* shows distribution-specific measures of a variable, such as skewness or inter-quantile-range, making variables comparable. This approach aims to trigger *hypothesis generation* by outlining tendencies between these variables.

Visual Analysis of Heterogenous Non-Image Data. Zhang et al. analyze subject groups in a web-based linked view system (Zhang et al., 2012). The resulting decision rules aim to categorize new subjects as they are added to the data. They define a cohort as parameter-divided subject group, differing from the epidemiological understanding of the term. The paper lacks details about handling of missing data, the definition of similarity or which statistical measures are applied. Generalized Pairs Plots (GPLOM' S) visualize heterogenous variables in a plot matrix (Emerson et al., 2013; Im et al., 2013). The plot depends on the type of variables, which are pairwise visualized. The plot matrix is well suited for an overview visualization, but takes up much screen space and is therefore only suitable for few variables at once (see Fig. 3). Dai and Gahegan defined a *Concept Map*, linking cancer-associated features by incorporating choropleth maps with bar charts, parallel coordinates and scatter plots with regression lines (Dai and Gahegan, 2005). The *Concept Map* is iteratively refined using primarily the underlying geographical data.

Decision Rule Driven Analysis of Medical Data. Related work in this field often focuses on clinical diagnosis. Closest to ours is the work of Glaßer (Glaßer et al., 2013) and Niemann et al. (Niemann et al., 2014), who both use decision rules for their analysis. Decision rules are easily readable and are well suited for classifying medical data. The results can be easily communicated to domain experts with little knowledge about the algorithm behind the results. The effort of analyzing the decision rules increases with decision rule complexity. Glaßer et al. use variables derived from DCE-MRI data capturing the perfusion in the tumorous tissue (Glaßer et al., 2013). They aim to

classify breast tumors. The extracted variables train a decision-tree classifier, concluding that the extracted features alone are not sufficient for tumor type classification. Niemann et al. assess hepatic steatosis (fatty liver disease) risk factors using decision trees (Niemann et al., 2014). They presented an interactive data mining tool, which can analyze association rules and highlight interesting relations, which may trigger new hypothesis about the data. We combine ideas of both papers by validating the significance of image-derived parameter aiming to derive new relationships by training decision trees to explain target variables.

Pinhero et al. applied association rule mining to create high-confidence decision rules based on variables such as age, gender and geographic location about the mortality rate in liver cancer data (Pinheiro et al., 2013). Sekhavat and Hoeber propose an interactive visualization technique for decision trees, where all decisions are displayed in a heat map (Sekhavat and Hoeber, 2013). The visualization translates the decision tree into a heat map, where each cell represents the connectivity between an initial value and an target value. The resulting interactive visualization is well suited for analyzing single, but not multiple decision trees.

SHIP Data Analysis. Klemm et al. analyzed the lumbar spine variability of of 490 SHIP-2 subjects (Klemm et al., 2013). They incorporate hierarchical agglomerative clustering to derive shape groups, yielding several groups of average shape and several outlier cluster. The extracted lumbar spine shape was correlated with subject size. The data extraction process is based on the detection algorithm and lumbar spine canal extraction presented in this paper. Clustering techniques applied to non-image data are strongly dependent on the chosen variables and distance measures.

Unique for our work is the combination of an IVA-approach by observing interesting feature relations in context of image-derived features using multiple decision trees for one visualization. We abstract the decision tree results similar to (Turkay et al., 2013), making them comparable in an overview visualization.

4 The Lumbar Spine Data Set

Our approach allows to analyze many variables simultaneously. Therefore, epidemiologists compiled the data set with a wide range of variables possibly correlating with lumbar back pain. The data set com-

prises of 6,753 subjects from two cohorts (4,420 from SHIP-Trend-0 and 2,333 from SHIP-2).

4.1 Non-Image Data

The variables range from somatometric variables describing body measures to medical examinations, such as laboratory tests as well as lifestyle factors as sporting activity or nutrition. Without the 9 image-derived parameter, the data set comprises of 134 variables:

- 21 metric variables, mostly describing somatometric variables and markers retrieved from blood analysis
- 113 categorical variables divided into
 - 43 dichotomous variables, mostly indicating the presence of a disease
 - 70 variables with more than two levels, indicating pain levels, nutrition and social factors, such as marital or retirement status

Many variables comprise missing data. Some variables are follow-up questions, covering the treatment of a disease. Others are exclusive to subgroups, such as women-specific variables, e.g. menstrual status. Sparse variables are statistically less resilient and need to be treated with care in the analysis.

4.2 Image Data

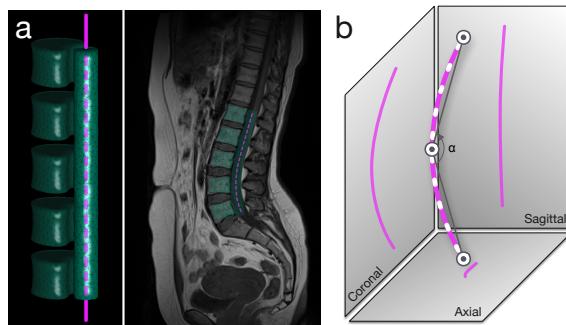


Figure 1: (a) left: Tetrahedron-based finite element model (FEM) of the lumbar spine, capturing the L1-L5 vertebrae as well as the lumbar spine canal. (a) right: FEM aligned to a MRI data set. The purple dashed line represents the centerline, describing the lumbar spine canal with 92 points. (b) Using the Frenet frame, we extract the weighted sum of curvature and torsion for all 92 points (white dashes). The curvature angle (α) is calculated using the top, bottom and middle point. We also extracted the variables for each projection axis to assess their information gain.

Four trained radiology technicians acquired the medical image data for each subject in a standardized way using a 1.5 Tesla MRI scanner (Magnetom

Avanto; Siemens Medical Solutions, Erlangen, Germany). The spine protocol consisted of a sagittal T1-weighted turbo-spin-echo sequence ($1.1 \times 1.1 \times 4.0 \text{ mm}$ voxels) and a sagittal T2-weighted turbo-spin-echo sequence ($1.1 \times 1.1 \times 4.0 \text{ mm}$ voxels) (Hegen-scheid et al., 2013).

A hierarchical finite element method was used to detect the lumbar spine in the MRI data (Rak et al., 2013). The tetrahedron-based finite element model is initialized using three user-placed landmarks. A click on the L3 vertebra center initialized the model, two clicks on the top and bottom of the vertebra describes an initial height estimation. The registered model captures information about the lumbar spine canal shape and the position of the L1-L5 vertebrae (Klemm et al., 2013) (Fig. 1 a). Each subject’s model is mapped to a common reference frame using procrustes analysis to compensate for differences in translation, rotation and scaling. The detection fails for several subjects due imprecise initialization, imaging artifacts or strongly deformed spines. 2,540 lumbar spine models were obtained and constitute the foundation of this work.

We have to assess the model accuracy to extract key figures from it. The detection model depicts the vertebrae positions, but lacks detailed information about their volume. It captures reliable information about spine canal curvature. The centerline representation of the lumbar spine canal was extracted from the detection model (Fig. 1 a) using the approach of (Klemm et al., 2013).

Using the Frenet frame, we calculated the following metrics from the model (Fig. 1 b) (Frenet, 1852):

- *Normalized Curvature* is calculated as weighted sum of curvature between all adjacent points describing the centerline: $\sum_{i=1}^I \frac{\text{curvature}_i}{i}$. We refer to total curvature from now on as *curvature*.
- *Normalized Torsion* (deviation of a curve from its current course) is calculated as weighted sum of torsion between all adjacent points describing the centerline: $\sum_{i=1}^I \frac{\text{torsion}_i}{i}$. We refer to total torsion from now on as *torsion*
- *Curvature angle α* is defined by the middle point of the spine canal centerline as *vertex* and the line between middle point and top/bottom point as *sides*.

These figures are also extracted in the sagittal, coronal and transversal projection of the model. We assess the information gain of each dimension using these projections. This gives us a total of 9 image-derived parameters.

5 EXPERIMENTS

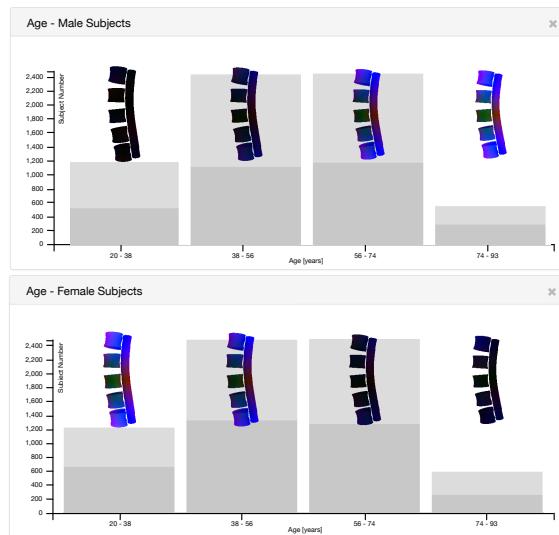


Figure 2: Correlation of age and gender regarding the lumbar spine size. The bar chart shows subjects divided into four different age groups (x-axis) and their subject number (y-axis). Each bar contains the mean lumbar spine of the respective group. The shape color encodes the distance to the overall male (top chart) or female (bottom chart) mean shape. The dark gray share of each bar encodes the portion of male (top chart) or female (bottom chart) subjects. Women have a higher life expectancy than men, hence their higher share in the old age group. Also, women are on average smaller than men, hence the larger shape similarity with older subjects. Due to bone erosion, older people are also smaller on average.

Spine shape is confounded by several somatometric variables. Larger people have also a longer spine and its shape is more straight. Since men are on average taller than women and people of old age shrink due to bone erosion, gender and age are also confounders (Fig. 2). Large body weight increases the spine load, resulting in a bend shape. All these variables need to be taken into account, when correlating spine curvature and torsion with non-image parameter. Since the gender confounder mainly encodes body-size, we decided to divide subjects into body-size groups. To avoid small outlier groups, epidemiologists recommend using quartiles to discretize metric variables (Klemm et al., 2014).

5.1 Preliminary Results

As first experiment we correlated the shape parameter with the binary back pain indicator using GPLOM's (Emerson et al., 2013). Since our image-derived variables are metric, their pairwise combinations are visualized using scatterplots. The combination of the

image variables with back pain is visualized as histogram at the left side of the matrix and as box plot on the right side. Figure 3 (right) shows the range of each variable as box plot. The projections to the transversal planes attract attention as they have many outliers. We conclude that curvature is not as reliable on the transversal plane as it is on the other planes.

5.1.1 Correlation Matrix

We calculated an association matrix to assess correlations between the image parameters. The pearson correlation coefficients between the numeric variables are depicted right of the matrix diagonal in Figure 3. *Curvature*, *curvature angle* and *torsion* correlate strongly with their planar projections. Also the mean curvature and the curvature angle correlate by a factor of -0.89 . Torsion does not correlate with any other image-derived variable.

5.1.2 Correlation of Image Parameters With Back Pain

Figure 3 shows the distribution of all image-parameters as generalized pairs plot. No statistically significant correlation could be observed through all subject groups. The box plots show no difference between subjects with and without back pain.

5.1.3 Assessing the Information Gain Using the PCA

To determine the information gain per image-derived variables, we calculated a PCA and compared the loadings per dimension. The first three principal components explain 75% of the variance in the image-derived variables. The first principal component explains 47% of the variance and weights primarily *mean curvature*, *curvature sagittal*, *curvature angle* and *curvature angle sagittal*. The second component, adding 16% of the variance, weights *curvature coronal* and *curvature angle coronal*. The third component explains 12% of the variance and weights *torsion* and *curvature transverse*. This supports our prior conclusion about the low information gain of the transverse planes. *Torsion* also adds little variance to the information space.

5.1.4 Heterogenous Correlations

We then expanded our focus on correlations of image-derived parameters with all other non-image variables. We applied a heterogenous correlation technique to derive correlations between all variables in the data set. The method uses the following correlation metrics for the different type combination:

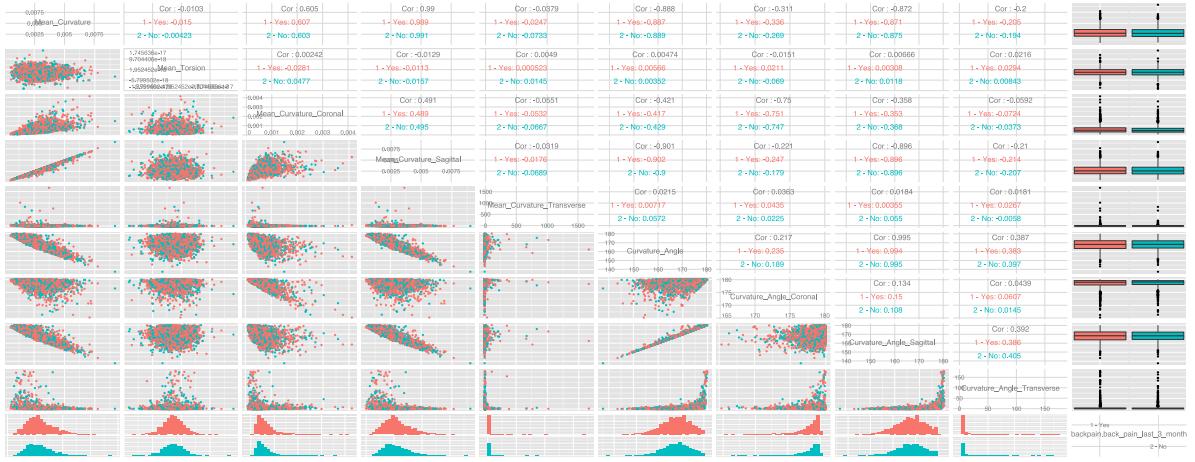


Figure 3: A generalized pairs plot of all image-derived variables colored by presence or absence of back pain. Pairwise combinations of image-derived variables are visualized using scatter plots left of the matrix diagonal and their correlation with back pain is denoted right of the matrix diagonal. The box plots (left) and histograms (bottom) display the distribution of each image-variable encoded with back pain. No correlation with back pain are present can be seen in this plot.

- Pearson product-moment for two continuous variables,
- Polyserial correlation for one continuous and one categorical variable, and
- Polychoric correlations for two categorical variables.

All correlation values are scaled between 0 – no correlation and 1 – identical. Some variables are too sparse for calculating correlations, for example *treatment of diabetes*, or *medication against high blood pressure*. We display the resulting *contingency matrix* using a heat map, mapping correlation values to color brightness with white for 0 and dark blue for 1 (Klemm et al., 2014). We calculated the contingency matrix for all size groups and looked for correlations between image- and non-image derived variables. The resulting contingency matrices are available as supplementary material and show no strong correlation with any of the parameter. Only weak correlations could be found for *Mean Curvature* with *gender* (0.42), *body size* (0.39) and *number of born children* (0.29). One surprising result was the small correlation with *torsion*, which correlated with almost no variables (p values between 0 and 0.05) and *parkinson* (0.24).

These observations brought us to the conclusion to incorporate more sophisticated data mining techniques to assess the influence of the image-derived parameters.

6 EVALUATION OF DECISION TREES

As described before, correlation coefficients fail to infer back pain status based on lumbar spine canal curvature and torsion. We rely on predictive classification trained to obtain a complex rule set on how combinations of the image-parameter explain non-image variables. Decision trees are a popular classification method in data mining for creating predictive models. Leafs of a decision tree represent class labels, branches represent feature conjunctions leading to the class labels. Decision trees are easy to understand and to read. They work with numerical as well as categorical data. This allows epidemiologists to interpret the results without having deep knowledge about the algorithm creating the tree. Readability is only granted for small trees, complex structures with many branches are not desirable. Too many branches also impose overfitting to the data (Mitchell, 1997).

6.1 The C4.5/C5.0 Algorithm

Using a trained data set, the C4.5 algorithm builds decision trees based on information entropy. Such a calculation requires a numeric or categorical target variable. The algorithm then tries to find a decision tree, which divides the samples using the input variables just like the target variables. This means, that every node in the tree represents the attribute which splits the data most efficiently into the target subset. The pseudocode for the algorithm is defined in Algorithm 1 (Kotsiantis et al., 2007). C5.0 is developed to

```

Check for base cases;
for each attribute  $a$  do
    | Find the normalized information gain ratio
    | from splitting on  $a$ ;
end
Let  $a_{best}$  be the attribute with the highest
normalized information gain;
Create a decision node that splits on  $a_{best}$ ;
Recurse on the sublists obtained by splitting on
 $a_{best}$ , and add those nodes as children of node;
Algorithm 1: Building a decision tree using the
C4.5 Algorithm

```

produce smaller decision trees than C4.5 and improve the execution time. We use the R implementation of C5.0 (Kuhn et al.,). Categorical attributes with more levels are biased with more information gain in a decision tree (Deng et al., 2011). Creating dummy variables bypasses this problem.

The actual use for the resulting tree is the classification of new observations (subjects). Yet, we are interested in the decision rules and the classification accuracy.

6.2 Interactive Display of Decision Trees

We have to create a decision tree for every non-image variable to analyze which one can be explained by image-derived variables. Since we have 134 non-image variables, the calculation yields a corresponding number of trees. Dividing the subjects into four groups using body size, the number increases to 402 trees. We have to abstract the results of the classification to keep the mental effort of interpreting the data low.

6.2.1 Visualization of Classification Results

We follow the Visual Analytics mantra by analyzing first, show the important and analyze further (Keim et al., 2008). A first analysis step was performed by applying the classification algorithm to the data. The optimal classification uses a few rules to precisely characterize the target variable. Therefore, we are interested in *small trees* with a *low classification error rate*. The two measures form the axis for a *scatter plot* of the classification results.

The Error Term. Normally the error rate for a classification is calculated with $\text{error} = \frac{\text{classifiedCorrectly}_n}{n}$ where n is the number of subjects. The metric usually works well for variables with uniform distribution. It distorts the result for other distribution types. If for

example a variable indicating a disease is negative for 90% of the subjects and the classifier simply assigns all subjects to *not ill* the error metric would yield an error of 10%, even though it is very bad. Our error term therefore incorporates the discriminative power of each manifestation and is denoted as follows:

$$\text{error} = 1 - \frac{\sum_{m=0}^M \frac{\text{classifiedCorrectly}_m}{\text{all}_m}}{\text{number}_{dimensions}} \quad (1)$$

M represents the set of manifestation of each variable. The error is scaled to denote perfect classification with 0 and 1 is equal to random selection. We consider a result rated below 0.25 a good classification. It allows for comparability of error rates between variables with different manifestation count.

Attribute Mapping The scatter plot axis are defined by tree size and the previously described error metric. This allows us to visualize a multitude of classification results in one plot. If we for example classify and compare the same variables for different subject groups, for example male and female subjects. We can support this by color coding the data points according to group affiliation.

Many follow-up variables are sparse, such as *medication of diabetes* or *reason of early retirement*. The classification algorithm may produce higher accuracy for variables with less subjects due to the small sample size. This makes these results less reliable. We removed variables with less than 100 subjects, since they have too few subjects to be statistically significant. Furthermore, we map the number of subjects associated with a variable to point diameter in the scatter plot. This allows to instantly assess the reliability of the result.

We apply a square root scale for the tree size axis to highlight data points with few decision rules. Outlier results with very large decision trees would otherwise distort the resulting plot.

6.2.2 Dummy Variables

Dummy variables convert a categorical variable with multiple manifestations into several binary variables. Each binary variables encodes the presence of a manifestation. For example, a pain indicator variable ranging from *1 - no pain* to *4 - large pain* is subdivided into four binary variables (*No pain - Yes/No* to *Large Pain - Yes/no*). One subject can only have one of these variables set to true. This is useful for our classification, because it allows to determine which manifestation of a variable can be described best using the image data parameter.

6.2.3 Interaction With the Visualization

The described visualization provides a good overview over the classification results. We still want to be able to display *details-on-demand* (Shneiderman, 1996) and examine a decision tree in detail. This is realized by clicking on an entry on the visualization, which then displays the corresponding decision tree in detail. This allows for sequentially analyzing the classifications.

6.2.4 Implementation

All analysis are carried out using R, a widely used programming language for statistical calculations and visualizations. The interactive visualizations are realized using the `ggvis`¹ package. As opposed to the standard R plots, `ggvis` allows to adjust visualization parameters using UI controls, such as sliders. In order to make the train of thought comprehensible, we used RMarkdown, which allows to create reports by combining R with the Markdown syntax. We used R Shiny² to make the report available as dynamic web application. It allows to combine both the ability to create comprehensible reports as well as UI-controls, to adjust visible data and parameters in the plot to allow the viewer to analyze the data. Furthermore, calculations based on a prior data selection can be redone within the report. The web-based approach allows us to quickly exchange results with our collaborating epidemiological partners. They can use the technique without installing any software. Exchanging the prototype becomes as easy as exchanging a hyperlink.

6.3 Results

We ran the analysis using different subject groups:

1. All subjects,
2. subdivision into *male* and *female*,
3. subdivision into *Body-Mass-Index quantiles* ($BMI = \frac{m}{l^2}$ where m is the *body mass* in kilogram and l is the *body size* in meter), yielding the groups (17, 24.7] (24.7, 27.4] (27.4, 30.5] (30.5, 48],
4. subdivision into *size quantiles*, yielding the groups (139, 164], (164, 171], (171, 177], (177, 202].

We plotted each group twice. The first plot shows all original variables. The second plot shows all categorical variables transformed into dichotomous dummy variables.

¹GGVIS Footnote

²R Shiny footnote

6.3.1 All Variables

The vast majority of parameters can not be described well using the classifier. This is reflected in the large amount of variables classified with an error rate above 0.6.

None of the pain indicators can be described reliably using the image-based parameters. The only variable reliably classified in this group is gender. It can be discriminated with an error rate of 0.31 using 7 rules and incorporates only curvature and curvature angle variables. The distinctness lies in the average difference in body size between male and female subjects. *Medication for high blood pressure* is classified for 1,058 subjects with an error rate of 0.47 solely based on coronal mean curvature. Almost all subjects who are medicated (796/1,058) were correctly classified, the vast majority of non-medicated subjects (262/1,058) are false-positive classified. Therefore the classifier is not as useful. The four body size groups could be characterized with an error rate 0.48, but the decision tree comprises 71 rules and imposes overfitting.

The analysis of the dummy variable yields similar results like the *blood pressure medication*. Variables, such as subjects sized 139–164 cm, between 64 and 90 years of age or nutrition related parameters are dominantly populated by one manifestation. The classifier neglects the other group and yields a error rate below 0.5.

6.3.2 Gender Groups

Classification using groups divided by gender do not produce satisfying results. Only *hypothyroidism* could be described for male subjects with an error rate of 0.24 for 110 subjects using the *mean curvature* and *curvature angle*. Since there are only 30 male subjects diagnosed with *hypothyroidism*, the statistical power of the result is reduced. The dummy variable analysis showed that female subjects of 139–164 cm body size could be discriminated using the *mean curvature* and *curvature angle*, with an error of 0.38.

6.3.3 Body-Mass-Index Groups

Gender could be classified for each *BMI* group using *mean curvature* and *curvature angle*. The accuracy varies between 0.31 (*BMI* of 30.5–48, 4 decision rules) to 0.39 (*BMI* of 24.7–27.4, 5 decision rules). The starting age of smoking could be characterized well with an error rate between 0.25 to 0.32 for all *BMI* groups except of subjects with *BMI* of 30.5–48. The result however is probably over fitted to the data due to tree sizes between 14 and 16. [Warum?](#)

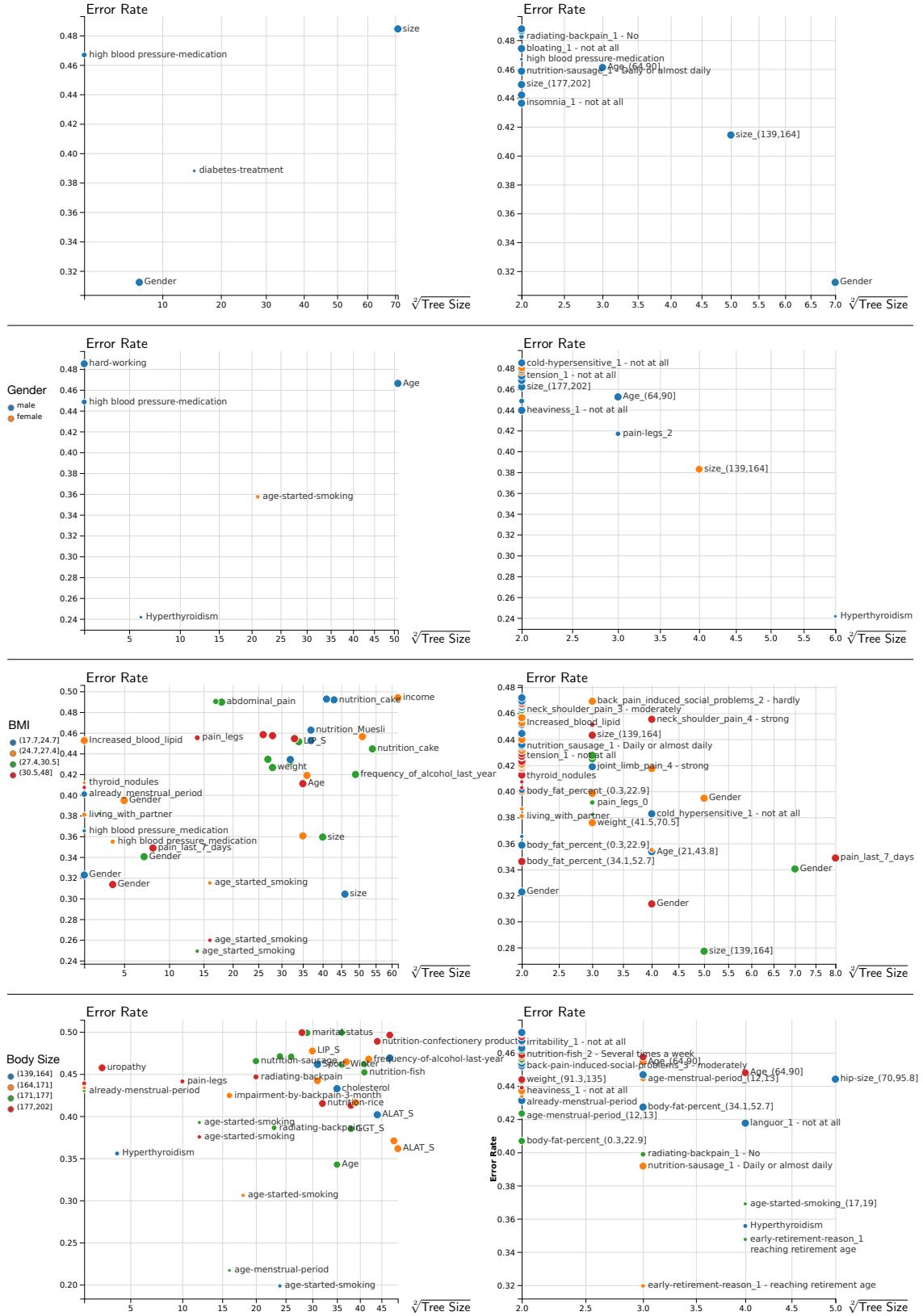


Figure 4: Scatterplot visualization of all C5.0 classification results. The x-axis shows the number of decisions of the underlying model, the y-axis the classification error rate (see Section 6.2.1). The left scatterplot shows the results for all variables, either metrics expressed via their quantiles, or categorical. The right scatterplot displays the dummy variables derived from the original variables. Group affiliation of a data point is color coded: no group (a), subdivision into *male* and *female* subjects (b), quartiles of *Body-Mass-Index (BMI)* (c) and quartiles of *body size* (d). The number of subjects represented in a variable is denoted using the dot diameter. We only display variables with an error rate below 0.5, results above this threshold are inaccurate. A interactive versions of the plot (see supplemental material) makes all data points clickable, showing the corresponding decision tree in a tool tip.

Some variables, such as *body size* can be described with an error rate of 0.3 to 0.36 but only using large decision trees with over 20 rules. Notable is also the increasing accuracy with increasing decision tree size. Using mostly *mean curvature* and *curvature angle*, the leg pain level can be predicted using 14 rules with an error rate of 0.46 for obese subjects (*BMI* higher than 30). The dichotomous variable whether the subject has felt pain in the last seven days can also be predicted for this group using the same features. The resulting tree consists of 8 rules and has an error of 0.35. Obese subjects are prone to back and leg pain due to a more stressed lumbar spine. How much the spine is deformed under the heavy stress seems to directly influence the pain levels for these subjects.

The dummy variable analysis shows many results using a decision tree with one rule based on *mean curvature* or *curvature angle* with accuracy between 0.35 and 0.47.

6.3.4 Size Groups

Many previously described results are confounded with subject size. Differences between *male* and *female* subjects are explained by the average *body size* difference. For example, large subjects are already characterized by their rather straight spine. The question is, whether the inter-group spine-variability parameter is enough for predicting other parameter or not. Dividing subjects into *body size* groups potentially highlights classifications not confounded by *body size*.

Large Decision Trees. Back pain-associated variables can be explained for various *size-groups*, but we could not extract universal rules. Radiating back pain could be described with error rate of 0.39 using 23 rules for subjects between 171 – 177 cm *body size* using *torsion* and *mean curvature*. For subjects sized 177 – 202 cm the accuracy drops to 0.47 using 20 decision rules. There are several decision trees for laboratory values, such as *alanine aminotransferase* value (relevant for diagnosis of liver or gallbladder illness) in the blood can be described with an error rate of 0.4 (139 – 164 cm) to 0.36 (164 – 171 cm). Similar values can be observed for *cholesterol* or *age*. Due to the large decision trees, these results are not usable and impose overfitting to the data.

Small Decision Trees. The dummy variables show several variables described using only one decision rule with accuracy between 0.42 to 0.47. Most of these variables have a dominant manifestation and the

classifier shows a low detection precision for the second manifestation. These variables include *nutrition* parameter, *thyroid disorder* and *social problems induced by back pain*.

7 CONCLUSION

The presented results indicate that the parameters *torsion*, *curvature* and *curvature angle* are not enough to characterize pathological changes in the lumbar spine. They are well suited to characterize subject gender due to their body size difference using *mean curvature* and *mean curvature angle*. This is due to the higher curvature of smaller subjects and the significant difference of average *body size* between *males* and *females*. Gender was characterized well in the analysis of all subjects as well as subjects divided by *BMI*.

Subjects grouped using *body size* have a similar spine shape and a lower variance. The remaining information within these groups is not enough to characterize back pain-related variables. For some *body size* groups, certain *nutrition* or *psychological* parameters can be predicted with a comparatively high error rate.

Another observation during the analysis is very large decision trees achieving low error scores. Their complexity impose overfitting to the data and they are not suitable for extracting universal rules.

7.1 Proposed Features

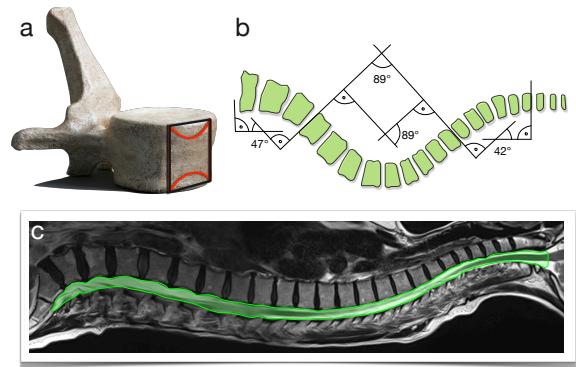


Figure 5: Proposed features. (a) dented vertebrae are a sign of heavy stress and bone absorption. (b, c) the spine canal shape can be used to characterize scoliosis (curvature sagittal) and lordosis (curvature coronal) as well as the *Cobb-Angles* (Cobb, 1948). (c) Spine canal thickness is associated with herniated disks.

Pathological deformation of the lumbar spine is usually the last stage and is associated with very strong

lower back pain [Referenz?](#). We have to capture earlier signs of pathological change in order to derive better predictive models.

A very early sign of pathological deformation is the *bone resorption in the center of a vertebra*. They change from being block-shaped to be dented in the center (Fig 5 a). This information can be obtained by segmenting the whole vertebra or the top and bottom point of each vertebra center. Another valuable variable would be the *spine canal thickness* (Fig 5 c). Low spine canal thickness can be an indicator for an impending herniated disk. Both surface texture of the vertebra and thickness of the spine canal are used to diagnose herniated disks. The *overall spine canal shape* is also of interest, since scoliosis and lordosis can be characterized more precisely by deriving the Cobb angles from this shape (Fig 5 b, c).

7.2 Applicability

Classification methods based on decision trees showed to be useful to assess the discriminative power of a variable set. Their ability to assess variable combination makes them more powerful than correlation coefficients calculated for each variable. This advantage comes with a much more complex output, making their results harder to assess and to abstract. Our method to plot derived metrics and custom tailored error measures proved to be effective. Huge result spaces could be navigated fast using our brushable visualization. Therefore the method is applicable not only for deriving information based on image data, but on all potential target variables.

Suggesting potentially interesting features when analyzing a condition is an important aspect of visual analytics in epidemiology (Klemm et al., 2014). We can achieve this also using the presented method by trying to describe several (dichotomous) target values with a decision tree constructed from all available data. This allows both to assess the discriminative power of the data set regarding the target variables, as well as the most important parameters for the classification.

7.3 Summary and Future Work

We showed, that *torsion*, *curvature* and *curvature angle* of the lumbar spine at the presented precision are not enough to characterize lumbar back pain in the SHIP data set. Our method made it possible to assess their discriminative power, which is largely limited for separating *male* and *female* subjects, *nutrition* variables, as well as different disease indicators. The C4.5 algorithm showed to be an effective tool for eval-

uating a set of derived metrics regarding their suitability to classify non-image parameters. Over-fitting to the data indicated by complex decision trees have to be taken into account into the analysis as well.

In our future work, we will focus on more precise models for extracting measures described in Section 7.1. As another focus, we want to include the method into existing visual analytics methods designed for analyzing shape information for epidemiological data (Klemm et al., 2014).

Combining the power of statistical analysis, visual analytics and advanced data mining techniques is essential for analyzing increasingly complex heterogeneous population data. These methods do not aim to replace the traditional epidemiological workflow, but rather intervenes at the weak points of standard statistical methods.

ACKNOWLEDGEMENTS

SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grant no. 03ZIK012), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania. Whole-body MR imaging was supported by a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg-Vorpommern. The University of Greifswald is a member of the ‘Centre of Knowledge Interchange’ program of the Siemens AG. This work was supported by the DFG Priority Program 1335: Scalable Visual Analytics.

REFERENCES

- Cobb, J. (1948). Outline for the study of scoliosis. *Instructional Course Lectures*, 5:261–275.
- Dai, X. and Gahegan, M. (2005). Visualization based approach for exploration of health data and risk factors. In *Proc. of the International Conference on GeoComputation. University of Michigan, USA*, volume 31.
- Deng, H., Runger, G., and Tuv, E. (2011). Bias of importance measures for multi-valued attributes and solutions. In *Artificial Neural Networks and Machine Learning-ICANN 2011*, pages 293–300. Springer.
- Emerson, J. W., Green, W. A., Schloerke, B., Crowley, J., Cook, D., Hofmann, H., and Wickham, H. (2013). The generalized pairs plot. *Journal of Computational and Graphical Statistics*, 22(1):79–91.
- Fletcher, R. H., Fletcher, S. W., and Fletcher, G. S. (2012). *Clinical epidemiology: the essentials*. Lippincott Williams & Wilkins.
- Frenet, F. (1852). Sur les courbes à double courbure. *Journal de Mathématiques Pures et Appliquées*, pages 437–447.
- Glaßer, S., Niemann, U., Preim, B., and Spiliopoulou, M. (2013). Can we Distinguish Between Benign and Malignant Breast Tumors in DCE-MRI by Studying a Tumors Most Suspect Region Only? In *Proc. of Symposium on Computer-Based Medical Systems (CBMS)*, pages 59–64.
- Hegenscheid, K., Seipel, R., Schmidt, C. O., Völzke, H., Kühn, J.-P., Biffar, R., Kroemer, H. K., Hosten, N., and Puls, R. (2013). Potentially relevant incidental findings on research whole-body MRI in the general adult population: frequencies and management. *European Radiology*, 23(3):816–826.
- Im, J.-F., McGuffin, M. J., and Leung, R. (2013). Gplom: The generalized plot matrix for visualizing multidimensional multivariate data. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2606–2614.
- Keim, D. A., Mansmann, F., Schneidewind, J., Thomas, J., and Ziegler, H. (2008). *Visual analytics: Scope and challenges*. Springer.
- Klemm, P., Lawonn, K., Rak, M., Preim, B., Tönnies, K., Hegenscheid, K., Völzke, H., and Oeltze, S. (2013). Visualization and Analysis of Lumbar Spine Canal Variability in Cohort Study Data. In *VMV 2013 - Vision, Modeling, Visualization*, pages 121–128.
- Klemm, P., Oeltze-Jafra, S., Lawonn, K., Hegenscheid, K., Völzke, H., and Preim, B. (2014). Interactive Visual Analysis of Image-Centric Cohort Study Data. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, page in print.
- Kotsiantis, S. B., Zaharakis, I., and Pintelas, P. (2007). Supervised machine learning: A review of classification techniques.
- Kuhn, M., Weston, S., and Coulter, N. C5.0 classification.
- Mitchell, T. M. (1997). Machine learning. 1997. *Burr Ridge, IL: McGraw Hill*, 45.
- Niemann, U., Völzke, H., Kühn, J.-P., and Spiliopoulou, M. (2014). Learning and inspecting classification rules from longitudinal epidemiological data to identify predictive features on hepatic steatosis. *Expert Systems with Applications*.
- Pinheiro, F., Kuo, M.-H., Thomo, A., and Barnett, J. (2013). Extracting association rules from liver cancer data using the fp-growth algorithm. In *Computational Advances in Bio and Medical Sciences (ICCABS), 2013 IEEE 3rd International Conference on*, pages 1–1.
- Rak, M., Engel, K., and Toennies, K. (2013). Closed-form hierarchical finite element models for part-based object detection. In *VMV 2013 - Vision, Modeling, Visualization*, pages 137–144.
- Sekhavat, Y. A. and Hoeber, O. (2013). Visualizing association rules using linked matrix, graph, and detail views. *International Journal of Intelligence Science*, 3:34.
- Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343. IEEE.
- Steenwijk, M., Milles, J., van Buchem, M., Reiber, J. H. C., and Botha, C. (2010). Integrated Visual Analysis for Heterogeneous Datasets in Cohort Studies. *Proc. of IEEE VisWeek Workshop on Visual Analytics in Health Care*.
- Turkay, C., Lundervold, A., Lundervold, A. J., and Hauser, H. (2013). Hypothesis generation by interactive visual exploration of heterogeneous medical data. In *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data*, pages 1–12. Springer.
- Völzke, H., Alte, D., Schmidt, C., et al. (2011). Cohort Profile: The Study of Health in Pomerania. *International Journal of Epidemiology*, 40(2):294–307.
- Zhang, Z., Gotz, D., and Perer, A. (2012). Interactive visual patient cohort analysis. In *Proc. of IEEE VisWeek Workshop on Visual Analytics in Health Care*.

APPENDIX

If any, the appendix should appear directly after the references without numbering, and not on a new page. To do so please use the following command:
\section*{APPENDIX}