

Interactive Visual Analysis of Lumbar Back Pain

What the Lumbar Spine Tells About Your Life

Blind

Keywords: Epidemiology, Interactive Visual Analysis, Classification, Multi Modal Data

Abstract: The goal of epidemiology is to provide insight into disease causations. Hence, subject groups (cohorts) are analyzed to correlate the subjects' varying lifestyles, their medical properties and diseases. Recently, these cohort studies comprise medical image data. We assess the predictive quality of image-derived variables for lower back pain. Therefore, an Interactive Visual Analysis (IVA) framework was created and tested with 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 variables 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 live healthier or to support the diagnosis of widespread diseases. So far, 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.

Cohort studies are an instrument of epidemiological research. Nowadays, 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 population aging (Hoy et al., 2010). 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, an

automatic classification approach for characterization of back pain based on lumbar back spine attributes is still missing. Furthermore, previous analyses mainly focused on a single variable (e.g. *BMI*) or a small subset of variables.

We present a thorough analysis of the image-derived data of a cohort study including lumbar spine datasets. We analyze their discriminative power for different back pain characteristics. For this purpose, we combine data mining algorithms with data visualization techniques. Then, Interactive Visual Analysis (IVA) 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 IVA framework, constructed to enhance the epidemiological workflow with iterative data driven techniques. 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,
- The detection of correlations between image-based and socio-demographic as well as medical variables for an improved hypothesis generation,
- The identification of the most important variables via data mining methods (including a novel semi-quantitative evaluation concept) embedded in our

IVA-framework

The analysis, as well as the presented IVA tool is available as supplementary material in a interactive R Markdown¹ document (blind.dnsalias.com).

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.
- Epidemiologists compile a list of variables depicting a hypothesis to prove 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 (Fletcher et al., 2012). For example, the effect of nutrition on prostate cancer requires age-normalization due to the increased incidence for older subjects. 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 Challenges of Epidemiological Data Analysis

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 influence on the human body. We focus on large scale cohort study data. These studies aim at collecting as much data as possible for each subject. The collected data can be analyzed regarding many diseases and conditions.

Epidemiological data are heterogenous and incomplete. For example, women-specific questions or data about a disease treatment only affects subjects suffering from this condition. Therefore, statistical analysis has to take missing data into account.

Epidemiological data acquisition includes a variety 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 necessary. For example, continuous data, such as age,

is often discretized into quantile bins. Every information reduction proposes an assumption about the data and can therefore introduce a bias. For example, 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. The analysis of these data is challenging. Individual diagnosis or manual segmentation of each body structure by radiologists is tedious, costly and comprises little reproducibility. Segmentation algorithms are not generally available and need to be carefully adapted 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

Lower back pain is one of the most frequent diseases in the western civilization (Hoy et al., 2010). Epidemiologists assume associations between lumbar back pain and lifestyle factors. These include nutrition, physical activity and body posture combined with physical stress at work. The exact causes as well as particularly vulnerable risk groups are unknown. 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 Medical Image and Non-Image Data (Klemm et al., 2014b) propose an extension to the previously described epidemiological workflow using IVA. 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 measures. We incorporate their IVA approach by concurrently displaying complex relationships between variables via decision trees. (Steenwijk et al., 2010) propose a coordinated linked view system for both image-related and non-image data. It incorporates multiple plot types, such as scatter plots, parallel coordinates and time plots with brushing and linking facilities. Similar to our work, they

¹Developed by RStudio, Inc; rmarkdown.rstudio.com

quantify image data and project it into the non-image data information space. (Turkay et al., 2013) follow a similar approach by deriving descriptive data metrics from image data. 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. A survey on image-centric cohort studies and strategies to analyze the resulting data is given in (Preim et al., 2014).

Visual Analysis of Heterogenous Non-Image Data. (Zhang et al., 2012) analyze subject groups in a web-based linked view system. The resulting decision rules aim to categorize new subjects as they are added to the data. They define a cohort as variable-divided subject group, differing from the epidemiological understanding of the term. The described method lacks details about handling of missing data, the definition of similarity or the choice of the statistical measures. Generalized Pairs Plots (GPLOM'S) visualize heterogeneous 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. GPLOM'S are well suited for an overview visualization, but take up much screen space and are therefore only suitable for few variables at once (see example in Fig. 3 later on). (Dai and Gahegan, 2005) define a *Concept Map*, linking cancer-associated features by incorporating choropleth maps with bar charts, parallel coordinates and scatter plots with regression lines. 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 et al., 2013) and (Niemann et al., 2014). Both use decision trees for their analyses. Decision trees are easily readable and are well suited for classifying medical data. The effort of analyzing decision trees increases with their complexity. (Glaßer et al., 2013) use variables derived from DCE-MRI data capturing the perfusion in the tumorous tissue. To classify breast tumors, they train a decision-tree classifier, concluding that the extracted kinetic and morphological features alone are not sufficient for tumor type classification. (Niemann et al., 2014) assess hepatic steatosis (fatty liver disease) risk factors using decision trees. They present an interactive data mining tool, which can analyze association rules and highlight interesting relations, which may trigger new hypothesis about the data. We combine both ideas by validating the significance of

image-derived variables. We aim at deriving new relationships by training decision trees to explain target variables.

(Pinheiro et al., 2013) 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. (Sekhavat and Hoeber, 2013) propose an interactive visualization technique for decision trees, where all decisions are displayed in a heat map. The visualization translates the decision tree into a heat map, where each cell represents the connectivity between an initial and a target value. The resulting interactive visualization is well suited for analyzing single, but not multiple decision trees.

SHIP Data Analysis. (Klemm et al., 2013) analyzed the lumbar spine variability of 490 SHIP-2 subjects. They incorporate hierarchical agglomerative clustering to derive shape groups, yielding several groups of average shape and several outlier clusters. The extracted lumbar spine shape was correlated with subject size. Our data extraction process is based on the detection algorithm and lumbar spine canal extraction presented by (Klemm et al., 2013). Clustering techniques applied to non-image data are strongly dependent on the chosen variables and distance measures (Klemm et al., 2014a).

Unique for our work is the combination of data mining techniques with an IVA-approach by observing interesting feature relations in the 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 the simultaneous analysis of many variables. Therefore, epidemiologists compiled the data set with a wide range of variables possibly correlating with lumbar back pain. The data set comprises 6,753 subjects from two cohorts (4,420 from SHIP-Trend-0 and 2,333 from SHIP-2) including non-image and image-derived data.

4.1 Non-Image Data

The non-image variables range from somatometric variables describing body measures to medical examinations, such as laboratory tests as well as lifestyle factors, e.g. sporting activity or nutrition. The data set

comprises of 134 variables (and 9 additional image-derived variables described in the next subsection):

- 21 metric variables, mostly describing somatometric variables and markers retrieved from blood analysis
- 113 categorical variables divided into
 - 43 dichotomous (binary) variables, mostly indicating the presence of a disease, e.g. pancreatitis or high blood pressure
 - 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 during the analysis.

4.2 Image-Derived Data

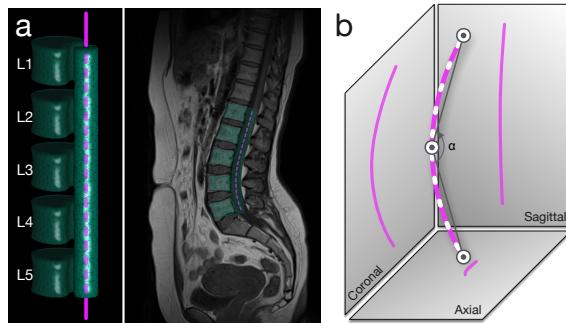


Figure 1: (a) left: Tetrahedron-based finite element model (FEM) of the lumbar spine, capturing the L1-L5 vertebrae as well as the lumbar spine canal. (a) right: FEM aligned to an MRI data set. The purple dashed line represents the centerline, describing the lumbar spine canal with 92 points. (b) Using 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 radiologists acquired the medical image data for each subject in a standardized way using a 1.5 Tesla Magnetic Resonance Imaging (MRI) scanner (Magnetom Avanto; Siemens Medical Solutions, Erlangen, Germany). The spine protocol consisted of a sagittal T_1 -weighted turbo-spin-echo sequence ($1.1 \times 1.1 \times 4.0 \text{ mm}$ voxels) and a sagittal T_2 -weighted turbo-spin-echo sequence ($1.1 \times 1.1 \times 4.0 \text{ mm}$ voxels) (Hegenscheid 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-defined landmarks. A click on the L3 vertebra center initializes the model, two clicks on the top and bottom of the vertebra describes an initial height estimation (Fig. 1 a). The registered model captures information about the lumbar spine canal shape and the position of the L1-L5 vertebrae. Each subjects' 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 to imprecise initialization, imaging artifacts or strongly deformed spines. 2,540 out of 6,753 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 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 metrics 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 yields a total of 9 image-derived variables. In the next section, we present the experiments we conducted to assess the influence of the lumbar spine shape to lower back pain.

5 EXPERIMENTS AND PRELIMINARY RESULTS

Spine shape is confounded by several somatometric variables. Larger people have also a longer spine with a straighter shape. Since men are on average

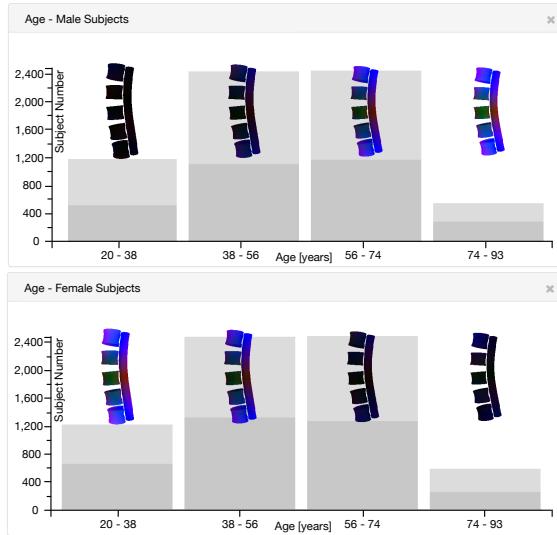


Figure 2: Correlation of age and gender regarding the lumbar spine size. The bar chart shows subjects divided into four different age groups (x-axis) and their subject number (y-axis). Each bar contains the mean lumbar spine of the respective group. The shape color encodes the distance to the overall 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.

taller than women and people of old age shrink due to bone erosion, gender and age are also confounders (Fig. 2). 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. Large body weight increases the spine load, resulting in a bend shape. All these variables need to be taken into account, when spine curvature and torsion is correlated with non-image variables. 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.

GPLOM Analysis. As a first experiment we correlated the shape variable with the dichotomous back pain indicator using GPLOM's (Emerson et al., 2013). Since our image-derived variables are metric, their pairwise combinations are visualized using scatter-plots. 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.

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

Correlation of Image Variables With Lumbar Back Pain. Figure 3 shows the distribution of all image-variables 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.

Assessing the Information Gain Using the Principal Component Analysis. To determine the information gain per image-derived variable, we calculated a Principal Component Analysis (PCA) and compared the dimension loadings. The PCA results in a linear projection of the variable space ordering their components by maximum variance. 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. The complete analysis can be found under the experiments section of the supplementary material (blind.dnsalias.com)

Heterogenous Correlations. We then expanded our focus on correlations of image-derived variables 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 combinations:

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

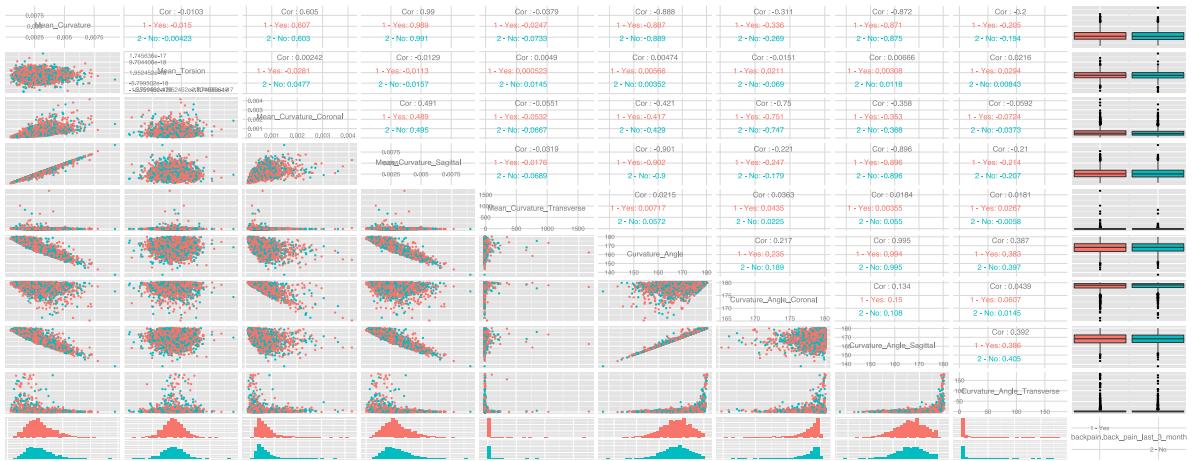


Figure 3: A generalized pairs plot of all image-derived variables colored by presence (red) or absence (turquoise) of back pain. Pairwise combinations of image-derived variables are visualized via scatter plots on the left of the matrix diagonal. Their correlation with back pain is denoted right of the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

All correlation values are scaled between 0 – no correlation and 1 – perfect correlation. Some variables are too sparse for calculating correlations, for example *treatment of diabetes*, or *medication against high blood pressure*. These variables have to be omitted in the analysis, since they are not statistically resilient. We display the resulting *contingency matrix* using a heat map, encoding correlation values with color brightness with white for 0 and dark blue for 1, as presented in (Klemm et al., 2014b). 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 variable (see blind.dnsalias.com). 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 of *torsion* with *parkinson* (0.24). Other than that, *torsion* correlated with almost no variables (p values between 0 and 0.05).

These observations brought us to the decision to incorporate sophisticated data mining techniques to assess the influence of the image-derived variables.

6 EVALUATION OF DECISION TREES

As described before, correlation coefficients fail to infer back pain status based on lumbar spine canal cur-

vature and torsion. We rely on predictive classification trained to obtain a complex rule set on how combinations of the image-variable explain non-image variables. Decision trees are a popular classification method in data mining to create 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. Epidemiologists can interpret the results without having deep knowledge about the classification algorithm. Readability is only granted for small trees. Too many branches also impose overfitting to the data (Mitchell, 1997).

6.1 The C4.5/C5.0 Algorithm

The C4.5 algorithm builds decision trees based on information entropy on a data set. Such a calculation requires a numeric or categorical target variable. The pseudocode for the algorithm is defined in Algorithm 1 (Kotsiantis et al., 2007). C5.0 is developed to produce smaller decision trees than C4.5 and improve the execution time. We use the R implementation of C5.0 (Kuhn et al., 2014). Categorical attributes with more levels are biased with more information gain in a decision tree (Deng et al., 2011). Creating a dummy variable by converting each manifestation of a categorical variable into a dichotomous variable bypasses this problem. In the following analysis, we strongly focus on the complexity of decision trees and the classification accuracy.

```

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

```

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 the same amount of trees. Further subdivision, e.g. by quantiles of *body size* increases the number to 402 trees. We have to abstract the results of the classification to keep the mental effort of interpreting the data low. This is included in our IVA framework.

6.2.1 Visualization of Classification Results

We follow the Visual Analytics mantra of 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*. The two measures form the axes for a scatter plot of the classification results. The scatter plot is our central element for the interactive analysis of decision trees.

The Error Term. Normally the error for a classification is calculated with $\text{error} = \frac{\text{correctlyClassified}_n}{\text{total}_n}$ 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. For example, if 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:

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

M represents the set of manifestation of each variable. The error is scaled to denote perfect classification with 0; 1 is equal to random selection. We consider a result rated 0.5 or below good enough to display it in the visualization, a value below 0.25 represents 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. Classification and comparison of variables for subject groups (e.g. male and female subjects) in one plot can be achieved by color coding group affiliation on the data points.

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. Therefore we provide a way to adjust the minimal number of subjects for each variable using a slider input. The initial value is empirically set to 100, marking a good tradeoff between sparse variables and statistical informative value. Furthermore, we map the number of subjects associated with a variable to point diameter in the scatter plot. This allows instant reliability assessment 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 dichotomous variables. Each dichotomous variable 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 dichotomous variables. 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 variable.

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 interactively displays the corresponding decision

tree in detail. This allows to sequentially analyze the classifications.

We also provide controls for adjusting the maximum classification error and minimum subject count for a variable. This gives the user control to abstract or refine the displayed information. Selecting a variable using a drop down menu allows the user to select the variable used for subject subdivisions, e.g. *gender* or *employment status*. Metric variables, such as *body size* are discretized using their quantiles. This allows to assess the influences of a variable to the classification process.

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*² package. As opposed to the standard R plots, *ggvis* allows to adjust visualization variables using user interface controls, such as sliders. In order to make the train of thought comprehensible, we used R Markdown, which allows to create reports by combining R with the Markdown syntax. We used R Shiny³ to make the report available as dynamic web application. It allows to combine the power of static R Markdown reports with dynamically parameterized *ggvis* plots. 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 experts. They can use the technique without installing any software. Exchanging the prototype becomes as easy as exchanging a hyperlink. The prototype is available under blind.dnsalias.com.

7 RESULTS

In this section, we show which non-image variables can be described using the 9 image derived variables. We also create subject groups to assess the influence of variables affecting the lumbar spine shape. We carry out the analysis using different subject groups:

1. All subjects,
2. subdivision into *males* and *females*,
3. subdivision by *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],

²Developed by RStudio, Inc; ggvis.rstudio.com

³Developed by RStudio, Inc; shiny.rstudio.com

4. subdivision by *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 (recall subsection 6.1).

7.1 All Variables

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

None of the pain indicators can be described reliably using the image-based variables. The only variable reliably classified in this group is *gender*. It can be discriminated with an error of 0.31 using 7 rules and incorporates only *curvature* and *curvature angle* related 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 of 0.47 solely based on *coronal mean curvature*. Almost all subjects who are medicated (796 of 1,058) were correctly classified, the vast majority of non-medicated subjects (262 of 1,058) are false-positive classified, yielding a poor quality of the classifier w.r.t epidemiological research. The four *body size* groups could be characterized with an error 0.48, but the decision tree comprises 71 rules and imposes overfitting.

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

7.2 Gender Groups

Classification using groups divided by *gender* do not produce satisfying results. Only *hypothyroidism* could be described for male subjects with an error 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.



Figure 4: Interactive 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 (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 below 0.5, results above this threshold are inaccurate. The interactive plot (see supplemental material) has clickable data points, displaying corresponding decision tree in a tool tip.

7.3 Body-Mass-Index Groups

Gender could be classified for each *BMI* group using *mean curvature* and *curvature angle*. The error 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 between 0.25 to 0.32 for all *BMI* groups except of subjects with *BMI* of 30.5 – 48. The result however is probably overfitted to the data due to tree sizes between 14 and 16.

Some variables, such as *body size* can be described with an error of 0.3 to 0.36 but only using large decision trees with over 20 rules. Using mostly *mean curvature* and *curvature angle*, the *leg pain level* can be predicted using 14 rules with an error 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. The stress induced spine deformation seems to influence the pain levels for these subjects directly.

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

7.4 Size Groups

Many previously described results are confounded by *subject size*. Differences between *male* and *female* subjects can be 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 variables 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 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 error increases to 0.47 using 20 decision rules. There are several decision trees for laboratory values, e.g. *alanine aminotransferase* value (relevant for diagnosis of liver or gallbladder illness) in the blood can be described with an error 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 an error 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*, *thyroid disorder* and *social problems induced by back pain*.

8 CONCLUSION

The presented results indicate that the image based variables *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* variables can be predicted with a comparatively high error.

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

Applicability. Classification methods based on decision trees have proven to be useful for assessing the discriminative power of a variable set. Their ability to consider variable combination makes them more powerful than correlation coefficients calculated for each variable. This advantage comes with a much more complex output, the results are more challenging 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 interactive 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. 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 variables for the classification.

Predictive Power of Image-Derived Variables. 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 evaluating a set of derived metrics regarding their suitability to classify non-image variables. Over-fitting to the data indicated by complex decision trees have to be taken into account as well.

Future Work. We provide a method for comparative analysis of decision trees independent of the variable manifestation count using interactive scatter plots. We applied the technique to derive insight into the predictive power of 9 image-derived variables for 134 non-image variables. The analysis was carried out for subjects divided by *gender*, *BMI*, and *body size* to assess their influence on the lumbar spine shape.

In our future work, we will focus on more precise models for extracting measures. Dented vertebrae are an early sign of a pathological deformation, so we want to capture this by segmenting the top and bottom point of each vertebra center. Another interesting metric is the spine canal thickness over the whole spine, capturing early signs of herniated disc disease. 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., 2014b).

Combining the power of statistical analysis, visual analytics and 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. Our method provides a novel way to gain insight into these complex data sets and amplifies hypotheses generation.

ACKNOWLEDGEMENTS

Omitted due to blind review.

REFERENCES

- 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.
- Hoy, D., Brooks, P., Blyth, F., and Buchbinder, R. (2010). The epidemiology of low back pain. *Best Practice and Research Clinical Rheumatology*, 24(6):769 – 781.
- 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., Frauenstein, L., Perlich, D., Hegenscheid, K., Völzke, H., and Preim, B. (2014a). Clustering Socio-demographic and Medical Attribute Data in Cohort Studies. In *Bildverarbeitung für die Medizin (BVM)*, pages 180–185.
- 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. (2014b). 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. (2014). 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.
- Preim, B., Klemm, P., Hauser, H., Hegenscheid, K., Oeltze, S., Toennies, K., and Völzke, H. (2014). *Visual Analytics of Image-Centric Cohort Studies in Epidemiology*, chapter Visualization in Medicine and Life Sciences III, page in print. Springer.
- 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*.