

Insight Into Lumbar Back Pain

What the Lumbar Spine Tells About Your Life

Paul Klemm¹, Sylvia Glaer¹, Kai Lawonn¹ Henry Vlzke² and Bernhard Preim¹

¹*Department of Simulation and Graphics, University of Magdeburg*

²*Greifswald*

{paul, sylvia}@isg.cs.uni-magdeburg.de, voelzke@anderAddy

Keywords: My Keywords in Title Case

Abstract: The abstract should summarize the contents of the paper and should contain at least 70 and at most 200 words. The text must be set to 9-point font size. Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

1 INTRODUCTION

Epidemiology is the study of causation of diseases. For this purpose, large population studies, such as the *Study of Health in Pomerania* (SHIP) [??] 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. These studies often comprise medical image data, e.g., tomographic image data sets, as well as general patient-specific attributes like gender, age and weight. Back pain is one of the most frequent diseases in the western civilization. Epidemiological analysis of back pain aim 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 yet. Furthermore, previous analyses mainly focused on a single parameter or a small subset of parameters [Quelle].

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. Therefore, 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. Da in der Auflistung oefter vom IVA-Framework die Rede ist: Wollen wir hierzu noch was schreiben, z.B. dass das im VIS-Paper definiert wurde und was wir darunter verstehen; sowie dass diese Arbeit als Erweiterung dieses Frameworks gedacht ist?

Our contributions are:

- An IVA-framework for back pain analysis based on image-derived variables of 2,240 subjects Ist diskutabel, so richtig interaktiv ist es nicht. IVA wurde auch noch nicht vorher definiert.
- 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 Wieder: IVA wurde noch nicht definiert

2 EPIDEMIOLOGICAL BACKGROUND

Epidemiology is the study of dissemination, causes and results of health-related states and events. 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 an *confounder* for prostate cancer. These *confounding* variables are often hard to find. Statistical correlation does not imply causation—epidemiologists need to assess the medical soundness of the statistical results.

2.1 Epidemiological Data

Epidemiological data can stem 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. 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. Women-specific question can only be acquired for female subjects, data about a disease treatment only for subjects suffering from this condition. Therefore, 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, since 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 part of the spine. 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 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

Our own work (VIS, VMV, BVM). [Sylvias paper with reference to the methods.](#) [More information necessary here!](#)

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 comprises 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 sport activity or nutrition.

4.2 Image Data

[Image Data](#) Four trained technicians acquired the medical image data for each subject in a standardized way using a 1.5 Tesla MRI scanner (Magnetom

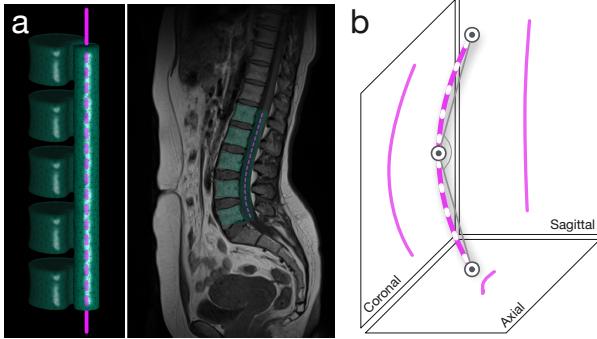


Figure 1: (a) left: Tetrahedron-based finite element model (FEM) of the lumbar spine, capturing the L1-L5 vertebrae as well as the lumbar spine canal. (a) right: FEM aligned to a MRI data set. The purple dashed line represents the centerline, describing the lumbar spine canal with 92 points. (b) Using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

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) (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-placed landmarks. A click on the L3 vertebra center gives the model the initial position, two clicks on the top and bottom of the vertebra describes an initial model 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). 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. In a previous work (Klemm et al., 2013; Klemm et al., 2014), we extracted a centerline representation of the lumbar spine canal from the detection model (Fig. 1 a). Using the Frenet formulas (Frenet, 1852), we extracted the following metrics from the model (Fig. 1 b):

- *Curvature* is calculated as weighted sum of curvature between all adjacent points describing the

$$\text{centerline: } \sum \frac{\text{curvature}_i}{\text{curvature}_{\text{all}}}.$$

- *Torsion* (deviation of a curve from its current course) is calculated as weighted sum of torsion between all adjacent points describing the centerline: $\sum \frac{\text{torsion}_i}{\text{torsion}_{\text{all}}}$.
- *Curvature angle* 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

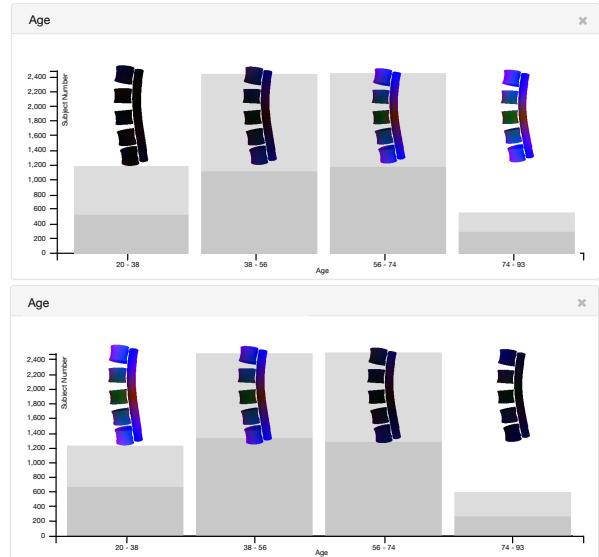


Figure 2: Correlation of age and gender regarding the lumbar spine size. The bar chart shows subjects divided into four different age groups (x-axis) and their subject number (y-axis). Each bar contains the mean lumbar spine of the respective group. The shape color encodes the distance to the overall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

and people of old age shrink due to bone erosion, gender and age are also confounder (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-height, 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 a generalized pairs plot (Emerson et al., 2013). Pairs plots visualize continuous and categorical variables pairwise in a matrix view. Scatterplots display pairs of continuous variables, bar charts for continuous variable as a function of a categorical variable and box plots pairs of categorical variables. 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 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](#). [We created-a generalized pairs plot for subject groups derived-by dividing them into quantile groups-of body-size](#). No statistical 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 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 variance, weighs *curvature coronal* and *curvature angle coronal*. The third component explains 12% of 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:

- *pearson product-moment* for two continuous variables,
- *polyserial* correlation for one continuous and one categorical variable, and
- *polychoric* correlations for two categorical variables.

[TODO: Abbildung einer Korrelationsmatrix? Waere sehr gross!](#) 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). This observation brought us to the conclusion to incorporate more sophisticated

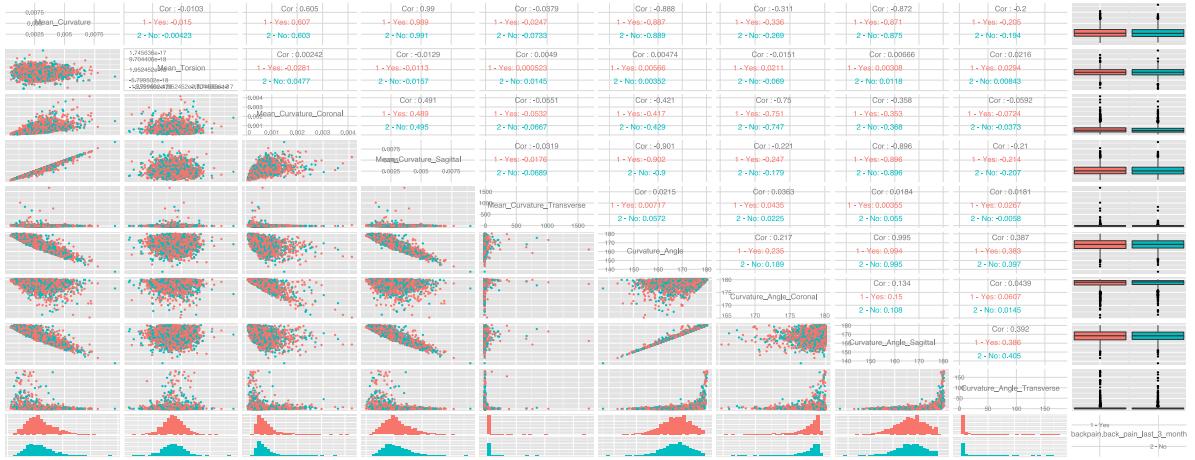


Figure 3: A generalized pairs plot of all image-derived variables (Emerson et al., 2013) colored by presence or absence of back pain. Pairwise combinations of image-derived variables are visualized using scatter plots left of the matrix diagonal and their correlation with back pain is denoted right of 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.

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 combination of the image-parameter explain non-image variables. Decision trees are a popular classification method in data mining for creating predictive models. Leaves 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 epidemiological domain experts 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 to the data. [Quelle Decision Tree](#)

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 vari-

ables just like the target variables. This means, that every node in the tree represents the attribute which splits the data most efficient 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 see 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. To make the results cognitive feasible, we have to abstract the results of the classification.

6.2.1 Visualization of Classification Results

We follow the visual information-seeking mantra and want to provide an overview first, then details-on-demand ?? . 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 $\text{error} = \frac{n_{\text{classifiedCorrectly}}}{n}$ 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 subject 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:

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

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. 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. Lets say we want to 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 to 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 distort otherwise the resulting plot.

6.2.2 Dummy Variables

Dummy variables abstract 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 however still want to be able to look display *details-on-demand* ?? and examine a decision tree in detail. This is realized by clicking on a entry on the visualization, which then displays the corresponding decision tree in detail (Fig. ?? b). This allows for sequentially analyzing the classifications.

6.2.4 Implementation

All analysis are carried out using R, a popular programming language for statistical calculations and visualizations. The interactive visualizations are realized using the ggvis¹ 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 make 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

¹GGVIS Footnote

²R Shiny footnote

within the report. The web-based approach allows us to quickly exchange results with our collaborating epidemiological partners. They can try out 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 *age quantiles*, yielding the groups (21, 43.8] (43.8, 54] (54, 64] (64, 90],
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.

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 a error rate above 0,6.

None of the pain indicators can be described reliably using the image-based parameter. 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. Surprisingly high is the classification of medication for high blood pressure. 1,058 subjects are classified 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 as the error-rate would indicate. The four body size groups could be characterized with an error rate 0,48, but the decision tree comprises of 71 rules and imposes overfitting.

The analysis of the dummy variable yields similar results like the blood pressure medication. Variables, such 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, reducing the statistical power of the result. The dummy variable analysis showed that female subjects of 139-164 cm body height could be discriminated using the mean curvature and curvature angle, with an error of 0,38.

6.3.3 Age Groups

Gender could be described for each age group using *mean curvature* and *curvature angle*. The accuracy varies between 0,35 (43,8 – 54 years of age, 3 decision rules) to 0,27 (21 – 43,8 years of age, 6 decision rules). The starting age of smoking could be characterized well with an accuracy between 0,25 to 0,3 for all age groups except of 21,1 – 43,8 years of age. The result however is probably over fitted to the data due to tree sizes between 16 and 23. [Feedback der Greifswalder notwendig, warum das so sein koennte.](#)

Many variables, such as body size can be described with an error rate of 0,35 to 0,45 but only using large decision trees with over 20 rules. Notable is also the increasing accuracy with increasing decision tree size. For subjects between 43,8 – 54 years of age, the classifier discriminated subjects with thyroid nodules with an error rate of 0,32.

The dummy variable analysis shows many results using a decision tree with one rule based on *mean curvature* with accuracy between 0,4 and 0,5. It shows that mean curvature can be used to predict dependencies to variables, such as *high blood pressure*, *hypothyroidism* or *nutrition*.

6.3.4 Size Groups

Many previously described results can be associated to be confounded with subject size. Differences in the gender analysis are mostly due to the average height difference between males and females. 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 height groups potentially highlights classifications not confounded by body-height.

Large Decision Trees. Back-pain associated variables can be explained for various size-groups, but we could not extract universal rules. Radiating back pain

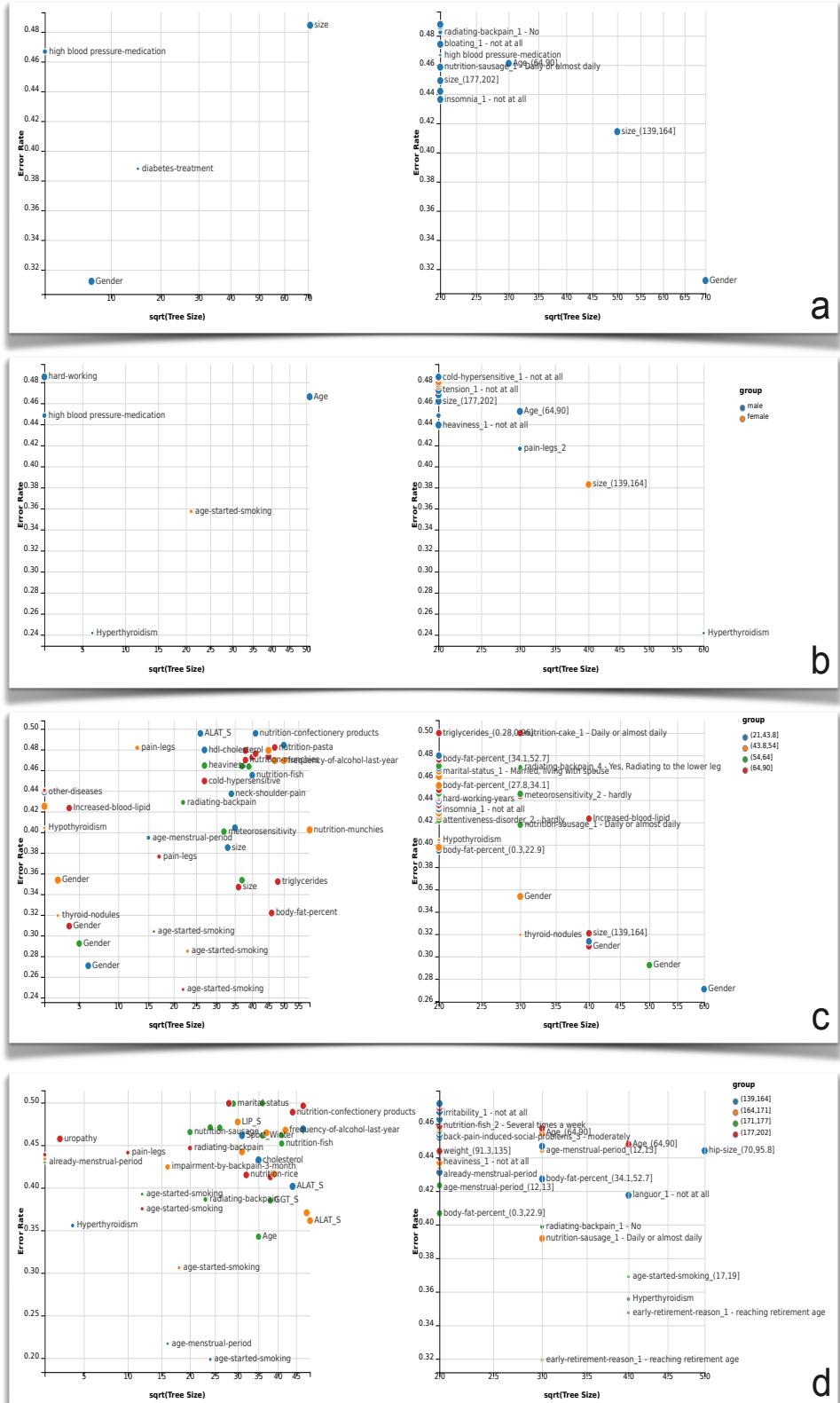


Figure 4: Scatterplot visualization of all C5.0 classification results. The x-axis shows the number of decisions of the underlying model, the y-axis the classification error rate (see Section 6.2.1). The left scatterplot shows the results for all variables, either metrics expressed via their quantiles, 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 age (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.

could be described with error rate of 0,39 using 23 rules for subjects between 171 – 177cm body size using torsion and mean-curvature. For subjects sized 177 – 202cm the accuracy drops to 0,47 using 20 decision rules. There are several decision trees for laboratory values, such as alanine aminotransferase value in the blood can be described with a high precision of 0,4 (139 – 164cm) to 0,36 (164 – 171cm). 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.

Auswertung der Anteile der Parameter an den Ergebnissen (Torsion, Curvature, etc. - welches ist der dominante Parameter?)

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 height difference using mean curvature and mean curvature angle. The reason for this lies in 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 age.

Subjects grouped using body height have a similar spine shape and a lower variance. The remaining information within these groups is not enough to characterize back pain related variables. Certain subgroups, however, can be classified for various size groups, such as nutrition or psychological parameters.

Another observation during the analysis are 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

Pathological deformation of the lumbar spine is usually the last stage and is associated with very strong lower back pain Zitat Frau Hegenscheid fragen. We



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

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 b). 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 c).

7.2 Applicability of the Method

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.

As described in a prior work (Klemm et al., 2014), the suggestion of 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 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 evaluating 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.
- 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 double courbure. *Journal de Mathmatiques Pures et Appliqués*, pages 437–447.
- 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.
- 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.
- 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.

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}`