Insight Into Lumbar Back Pain
What the Lumbar Spine Tells About Your Life

Paul Klemm[1], Sylvia Glaer[1], Kai Lawonn[1] Henry Vlzke[2] and Bernhard Preim[1]
[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.

1 INTRODUCTION

Epidemiology is the study of causation of diseases. 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, helping people to make their lifestyle healthier or helping in diagnosing a disease. Epidemiological research is strongly hypothesis driven. Observations made by clinicians are translated into hypothesis, which are then statistically evaluated using data variables from epidemiological studies.

Modern cohort studies often comprise medical image data. Short summary

Back pain is one of the most frequent diseases in the western civilization.

Our goal is to combine data mining algorithm with data visualization to provide insight into the quality of image derived data to analyze if it acts as a risk factor for a disease.

More than 5 variables are rarely analyzed simultaneously. Our contributions are:

- Analyzing back pain using image-derived variables of 2,240 subjects.
- Assessing the suitability of lumbar spine shape for diagnosing back pain
- Analyzing correlations between between image-based and socio-demographic as well as medical parameters.
- Identification of most important variables using data mining methods noch schoen schreiben
- semiquantitative Auswertmglichkeiten werden dem User oft zur verfgung gestellt
- Techniken sind als Teil des IVA-Frameworkes zu verstehen

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 (**?**):

- 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

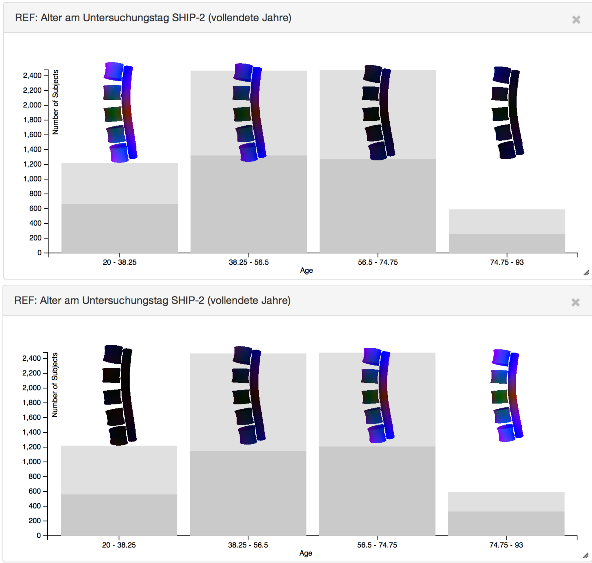

Figure 1: Age-Gender

Our approach allows to analyze many variables simultaneously. Therefore, epidemiologists compiled the data set with a 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

From VIS'14 Paper The lumbar spine was detected in the image data using a hierarchical finite element method by Rak et al. [32]. This semiautomatic method requires the user to initialize the tetrahedron-based finite element models (FEM) with a click on the L3 vertebra. Two user-defined landmarks on the top and bottom of the L3 vertebra describe an initial

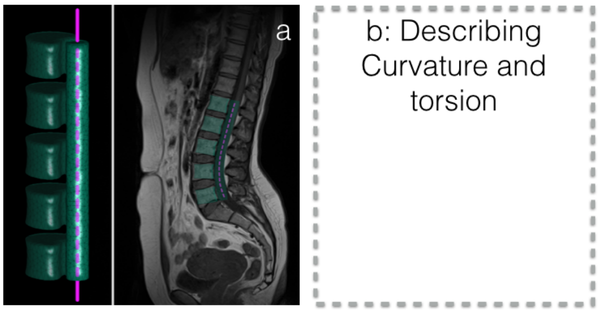

Figure 2: Centerline

model height estimation. The model uses a weighted sum of T1and T2-weighted MR images to detect the lumbar spine shape. Once registered, it captures information about the shape of the lumbar spine canal as well as the position of the L1-L5 vertebrae [21]. Due to incorrect initialization, strongly deformed spines, contrast differences and artifacts, the model was not able to detect lumbar spines for all subjects. We obtained and worked with 2,540 tetrahedron models of the lumbar spine. For clustering, we extracted the centerline of the lumbar spine canal, which captures information about lordosis and scoliosis (the medical terms for spine curvature) [21].

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 (**?**; **?**), we extracted a centerline representation of the lumbar spine canal from the detection model (Fig. 2 (a)). Using the Frenet formulas (**?**), we extracted the following metrics from the model (Fig. 2 (b)):

- *Curvature* is calculated as weighted sum of curvature between all adjacent points describing the centerline: $\sum \frac{curvature_i}{curvature_{a}ll}$.
- *Torsion* (deviation of a curve from the course) is calculated as weighted sum of torsion between all adjacent points describing the centerline: $\sum \frac{torsion_i}{torsion_{a}ll}$.

These figures are also extracted in the sagital, coronal and transversal projection of the model. We assess the information gain of each dimension using these projections. This gives us a total of 8 image-derived parameters.

5 EXPERIMENTS

5.1 Preliminary Results

- Direkter Vergleich
- Verteilung des Means
- Korrelationsmatrix
- PCA

Weil das nicht ausreichend war, haben wir einen C5.0 Baum erstellt, um alle potentiellen Zusammenhnge zwischen bildbezogenen und Nichtbildbezogenen Parametern herzustellen - Was kann man mit den Bilddaten erklren und was nicht?

5.2 Experimental Settings and Results

Decision Tree erklaeren, ebenfalls Umrechnen in die Dummy-Variablen. Was sind die Zielvariablen

- Altersgruppierung
- Male/Female
- Gewicht (BMI - dnn, normal, dick)

6 EVALUATION OF DECISION TREES

- Vorgehen mit Scatterplot erklren
- Beschreiben der Ergebnisse
- Beschreiben des interaktiven Systems zum Anzeigen der Decision Trees

We can discriminate back pain using non-image data, but not with image-derived parameter Welche parameter knnen wir gut unterscheiden

7 CONCLUSION

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

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*}