

Supporting an Early Detection of Diabetic Neuropathy by Visual Analytics

M. Luboschik¹, M. Röhlig¹, G. Kundt², O. Stachs³, S. Peschel³, A. Zhivov³, R.F. Guthoff³, K. Winter⁴, H. Schumann¹

¹Institute of Computer Science, ²Institute for Biostatistics and Informatics in Medicine and Ageing Research,

³Department of Ophthalmology, University of Rostock, Germany.

⁴Translational Centre for Regenerative Medicine (TRM), University of Leipzig, Germany.

Abstract

In this paper, we describe a step-wise approach to utilize ophthalmic markers for detecting early diabetic neuropathy (DN), the most common long-term complication of diabetes mellitus. Our approach is based on the Visual Analytics Mantra: First, we statistically analyze the data to identify those variables that separate DN patients from a control group. Afterwards, we show the important separating variables individually, but also in the context of all variables regarding a pre-defined classification. By doing so, we support the understanding of the categorization in respect of the value distribution of variables. This allows for zooming, filtering and further analysis like deleting non-relevant variables that do not contribute to the definition of markers as well as deleting data records with false data values or false classifications. Finally, outliers are observed and investigated in detail. So, a third group of potential DN patients can be introduced. In this way, the detection of early DN can be effectively supported.

Categories and Subject Descriptors (according to ACM CCS): I.3.8 [Computer Graphics]: Applications—, J.3 [Computer Applications]: Life and Medical Sciences—, H.5.0 [Information Interfaces and Presentation]: General—

1. Introduction

In 2006, diabetes mellitus was the first non-communicable disease to receive a United Nations General Assembly Resolution (61/225), recognizing it as a chronic, debilitating and costly disease associated with major complications that pose severe risks to the entire world. A serious complication is diabetic neuropathy (DN), which currently possesses no validated marker for its early detection. However, there are advances in the development of ophthalmic markers of DN, still lacking of statistical reliable methods for quantifying as well as for visualizing them. In fact, an adequate quantification and visualization strategy will provide a useful non-invasive mean for evaluation and detection of early DN.

In this paper we describe our approach to visually analyze the data of two clinical studies in order to identify particular characteristics describing DN patients. We aim at a mix of computational and visual interactive methods. By means of visual exploration we want to support the understanding of the data, and by applying statistics we want to achieve quantified results which are needed to come up with robust findings. The remainder of the paper is organized as follows: Section 2 describes the clinical background as well as the

data of the clinical studies. Section 3 describes our solution aiming at identifying specific and sensitive markers for DN. In tight cooperation with domain experts from statistics and ophthalmology, we developed a step-wise approach with regard to the Visual Analytics Mantra. For each step, we select appropriate existing techniques combining them into an integrated workflow. Section 4 concludes our discussion with an outlook on further work.

2. Clinical Background

In 2011, the UN has convened a high-level meeting to prevent and control non-communicable diseases, diabetes leads the agenda, as its worldwide prevalence predictions for 2025 have already been exceeded. The most common long-term complication of diabetes mellitus is diabetic neuropathy (DN). Population-based studies reported prevalence for DN ranging from 8 to 54% in type 1 and from 13 to 46% in type 2 diabetes [ZRD*08]. Common symptoms of DN occur usually in the feet or legs and include pain, numbness and burning. DN is the primary risk factor for foot ulceration which can lead to amputation [KK09]. DN can also affect the autonomic nervous system and predicts mortality. One impor-

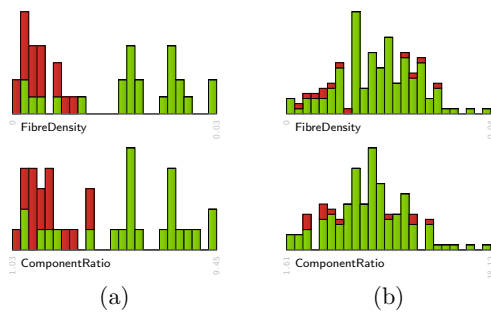


Figure 1: Stacked histograms demonstrate the value distribution of DN patients (red) and control group (green): (a) Rostock study, (b) Düsseldorf study.

tant issue is the detection of the earliest stage of diabetes and pre-diabetes at which neuropathy starts to develop. However, there is no validated marker for the detection of early DN. Thus, it is commonly diagnosed at advanced stages when significant nerve dysfunction and damage have already developed which are not reversible. In fact, once DN has become clinically evident, the decline in nerve function, as assessed by nerve conduction studies and clinical examination, progresses linearly despite good glycaemic control.

Therefore, to effectively prevent the clinical endpoints of DN, it is essential to detect nerve fibre abnormalities as early as possible using specific and sensitive markers at a sub-clinical stage of the disease, allowing effective approaches for intervention before irreversible damage. Reliable and clinically meaningful markers for DN could also be used to monitor early progression of the disease. There exist ophthalmic markers of DN that have been used to demonstrate that DN is associated especially with morphological degradation of corneal nerves [Efr11, ZPZ*14]. In vivo confocal laser scanning microscopy (CLSM) of the cornea – developed by the authors in [AZE*11] and currently established in more than 300 clinical/experimental sites worldwide – is able to identify early nerve fibre damage in the corneal sub-basal nerve plexus (SBP). The SBP image analysis is done automatically using the approach in [WZG*10] and is based on morphological (length, diameter, density) and topological (continuity and connectivity) features. This automated quantification strategy accelerates the use of in vivo CLSM in clinical practice, making it a useful noninvasive tool for evaluation of small fibre neuropathy. However, there is a lack of statistical reliable methods of nerve fibre quantification including tailored visualizations tools. To identify specific and sensitive markers that allow for detecting nerve fibre abnormalities as early as possible we examined the data of two clinical studies: (1) of longtime diabetic patients [ZWH*13] and (2) recently diagnosed type 2 diabetes patients [ZPZ*14], each including matched controls.



Figure 2: TTPC-Table Lens images [JTS08] show the raw data of the studies.

3. Our Approach

The general aim is to support the finding of specific DN markers recorded by CLSM to replace existing unreliable classification scores or invasive methods like skin biopsy. A significant step towards this goal is to identify those variables of the clinical studies that separate DN patients from people of the control group. To supporting this aim, we developed the tool EyeVis that provides computational and visual interactive techniques. Basically our tool offers the following functionality:

- Analytic techniques
 - Discriminant analysis as a preprocess
 - Clustering (SOM and hierarchical)
 - Computation of cluster separability (e.g., [SNLH09])
- Visual techniques
 - TTPC-Table Lens [JTS08] for overview images
 - SPLOM and Matrix views enhanced with color-coded background for showing measures or classifications
 - Parallel Coordinates enhanced with Stacked histograms [PMBS14] to analyze particular variables
- Interaction Techniques
 - Selecting, Filtering, Linking & Brushing, re-ordering

For applying the tool EyeVis, we suggest a step-wise approach with regard to the Visual Analytics Mantra: Analyse First - Show the Important - Zoom, Filter and Analyse Further - Details on Demand [KMS*08].

1) Analyze first: Discriminant analysis is applied to identify a linear combination of quantitative predictor variables that best characterizes the differences between DN patients and controls. Then the resulting discriminant function was used as a linear classifier for reclassification, for classification of new cases with measurements for the

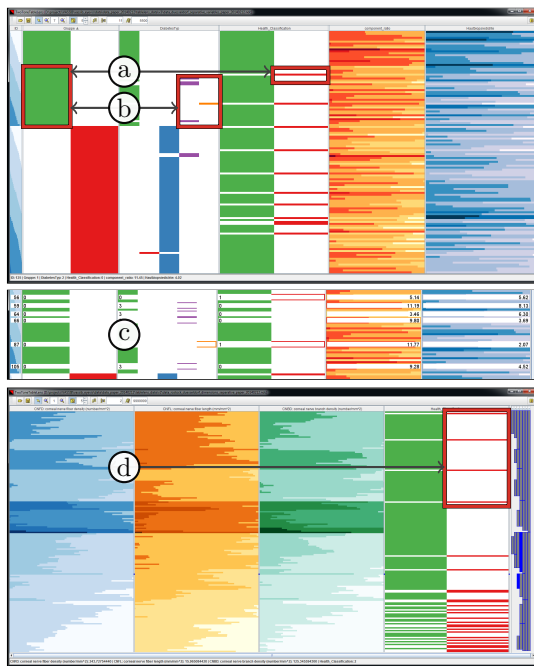


Figure 3: Analyzing data quality: identifying a) non-diabetic neuropathy cases, b) false classifications (non-diabetic patients associated with diabetes type 1 or 2) and c) zooming in for details of those. Remaining outliers are shown in d).

predictor variables but unknown group membership and additionally for dimensionality reduction.

Considering the Rostock study, all variables are included and the multivariate test by Wilks' Lambda results in $p < 0.001$ and indicates a highly significant difference between the two groups centroids (the means of all variables simultaneously). Obviously, the discriminant function including all variables works very well. Overall 95% (38 out of 40) of the patients from Rostock are assigned to the correct group. Within the healthy group 90% (18 out of 20) are classified correctly but 100% in the group of DN. In a second step we used the procedure to explore how many and which variables among all are most useful and sufficient for discriminating among the groups. For building a model in a stepwise manner – entering or removing one predictor variable from the model at each step – for identifying a useful subset of predictor variables decision was made by using the change in Wilks' Lambda. As result, only the variables "Percentage of covering of nerve fibres in respect to image area" and "Nerve fibre density" were of relevance for sufficient separation between the two groups. Discriminant function including only these two variables has the same power as the full model. Again, 95% (38 out of 40) of the Rostock patients are assigned to the correct group.

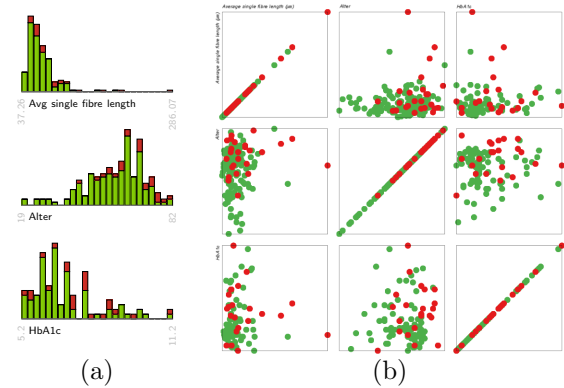


Figure 4: Example of 3 non-separating variables: a) shown by stacked histograms and b) by SPLOM.

However, regarding the study of patients from German Diabetes Center, University Clinics of Düsseldorf results were less conclusive. Again, only two variables ("Percentage of covering of nerve fibres in respect to the image area" and "density in skin biopsy" or "Percentage of covering of nerve fibres in respect to the image area" and "Normalized number of endpoints" in a model without "density in skin biopsy" (due to its invasive character)) were of relevance for sufficient separation. Then again discriminant function including only two variables has the same power as the full model. But the reclassification rates were clearly inferior to the data from Rostock. About only two third correct assignments would be recorded in conflict to 95% of Rostock data. Assessment of data of individual cases with extremely misclassification was undertaken and continues.

Resulting from this analysis step we have a first hypothesis about 2 separating CLSM variables. It can be shown that these variables in fact subdivide both groups with regard to the Rostock study and fail with regard to the Düsseldorf study. Thus, further investigations are necessary to determine specific and sensitive DN markers.

2) Show the Important: In the next step we use visualization to support an understanding of the statistic results. Figure 1 communicates the findings by stacked histograms. Figure 1a shows that DN patients (red) have lower values for both separating variables. In contrast, figure 1b does not show such a clear value distribution. This dissimilarity illustrates the different results of the computation step for both studies.

Now, we explore each data set according to all available variables to find alternative separating variable combinations. For this purpose we apply the TTPC-Table Lens as suggested in [JTS08]. This visualization technique is an extension of the well-known Table Lens in two regards: (1) the original Table Lens bars are replaced by TTPC

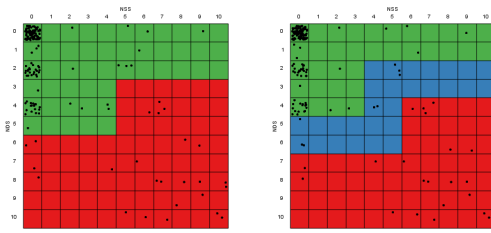


Figure 5: Matrices showing the data in relation to the unreliable DN classifications NDS and NSS: a) color-coded background in respect of the two groups (red: DN patients, green: control group) and b) with the additional fuzzy group (blue).

value representations [SMY*05], i.e. darker and longer bars represent higher values, and (2) the rows of the table are ordered by a two-step procedure; SOM clustering and hierarchical clustering. In this way, correlating variables can be easily perceived by similar value distributions. In the same way, outliers clearly stick out. Figure 2 shows the data of both studies by the TTPC-Table Lens. DN patients of the Rostock study (marked by the rectangle) have lower values in regard to several variables compared to the control group, whereas this does not hold for the Düsseldorf study. Since we need markers, which can be applied generally (and not only for one particular data set), further examinations are needed.

3) Zoom, Filter and Analyze further: The next step is data fusion. All variables that are given only in one of the studies are filtered out, and then the data sets are merged. Afterwards, an interactive analysis comes into play. Data records are checked to identify false values and classifications, and to filter the data accordingly (see figure 3). Variables are checked whether they contribute to separate DN patients. Variables with similar value distributions for both groups – DN and control – are deleted (see figure 4). The result of filtering is a cleansed data set. However, re-clustering the data (see step 2) with reduced variables and cases still reveals outliers (figure 3d), i.e. both groups are not sufficiently separated.

4) Details on demand: Finally, we explore the remaining outliers in more detail. Figure 5a shows the patients in regard to an *unreliable* DN classification. The data records are represented by dots, and the background is color-coded in respect of the classification. Most of the dots of DN patients (red) can be seen in the lower right corner, whereas most of the dots of the control group are placed in the upper left corner. The outliers (as marked in figure 3d) are located along the classification boundary. This motivates us to introduce a third class with people *potentially* having DN. By interactively specifying such a fuzzy region (blue region in figure 5b) the separation of the both groups, DN patients and control group, is significantly improved. To be more precise, the majority of DN

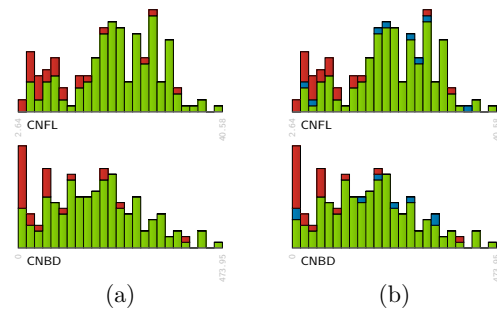


Figure 6: The stacked histograms show variables characterizing DN patients by low values: a) two groups with outliers (red: DN patients, green: control group) and b) with the additional fuzzy group (blue) eliminating almost all red outliers.

patients (red) are now located in the low value ranges of the stacked histograms while simultaneously all but two outliers are assigned to the fuzzy group (blue) (figure 6b).

The result of the described visual analysis process is a deeper insight into the classified data. The statistic part provides two separating variables. However, these variables do not sufficiently separate the data in any case. The interactive visual part helps with explaining this result, and supports the definition of a fuzzy group.

Based on this, new and unclassified data that will be gathered in future can be examined: If a new data record shows the same value distribution as the red group, we would assume DN. With a value distribution as the green group, we do not assume DN for now. If the new record is an outlier in one of both groups, or if it belongs to the fuzzy group, further investigations will be necessary to decide about DN. In this way, the detection of early DN can be effectively supported.

4. Concluding Remarks

Recently, CLSM has emerged as a promising non-invasive technique for the detection of small nerve fiber alterations in diabetic patients. Our team of experts from different domains (statistics, ophthalmology, visual analytics) developed a workflow tailored to the needs of identifying and evaluating markers for detecting DN as early as possible. Such markers are useful to expose nerve fibre abnormalities allowing for effective approaches for intervention before irreversible damage. Our investigations are based on classified data of two clinical studies. Since extracting the data from images is attended by different uncertainties, our future work will encompass appropriate interactive visual means to support the data extraction by integrating it into our workflow.

Future work also includes the development of an easy-to-use interface for clinical use to evaluate the markers practically and in this way to come up with hypotheses about DN.

References

- [AZE*11] ALLGEIER S., ZHIVOV A., EBERLE F., KOEHLER B., MAIER S., BRETTHAUER G., GUTHOFF R. F., STACHS O.: Image reconstruction of the subbasal nerve plexus with in vivo confocal microscopy. *Investigative Ophthalmology & Visual Science* 52, 9 (2011), 5022–5028. doi:10.1167/iovs.10-6065. 2
- [Efr11] EFRON N.: The glenn a. fry award lecture 2010: Ophthalmic markers of diabetic neuropathy. *Optometry and Vision Science* 88, 6 (2011), 661–683. doi:10.1097/OPX.0b013e3182171020. 2
- [JTS08] JOHN M., TOMINSKI C., SCHUMANN H.: Visual and analytical extensions for the table lens. In *Proceedings of Visualization and Data Analysis (VDA)* (2008), SPIE/IS&T, pp. 680907–1–680907–12. doi:10.1117/12.766440. 2, 3
- [KK09] KUMAR A. R., KAPLOWITZ P. B.: Patient age, race and the type of diabetes have an impact on the presenting symptoms, latency before diagnosis and laboratory abnormalities at time of diagnosis of diabetes mellitus in children. *Journal of Clinical Research in Pediatric Endocrinology* 1, 5 (2009), 227–232. doi:10.4274/jcrpe.v1i5.227. 1
- [KMS*08] KEIM D. A., MANSMANN F., SCHNEIDEWIND J., THOMAS J., ZIEGLER H.: Visual analytics: Scope and challenges. In *Visual Data Mining*, Simoff S. J., Böhlen M. H., Mazeika A., (Eds.). Springer-Verlag, 2008, pp. 76–90. doi:10.1007/978-3-540-71080-6_6. 2
- [PMBS14] PHAM T., METOYER R., BEZRUKOVA K., SPELL C.: Visualization of cluster structure and separation in multivariate mixed data: A case study of diversity faultlines in work teams. *Computers & Graphics* 38, 1 (2014), 117–130. doi:10.1016/j.cag.2013.10.009. 2
- [SMY*05] SAITO T., MIYAMURA H. N., YAMAMOTO M., SAITO H., HOSHIYA Y., KASEDA T.: Two-tone pseudo coloring: Compact visualization for one-dimensional data. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis'05)* (2005), pp. 173–180. doi:10.1109/INFVIS.2005.1532144. 4
- [SNLH09] SIPS M., NEUBERT B., LEWIS J. P., HANRAHAN P.: Selecting good views of high-dimensional data using class consistency. *Computer Graphics Forum* 28, 3 (2009), 831–838. doi:10.1111/j.1467-8659.2009.01467.x. 2
- [WZG*10] WINTER K., ZHIVOV A., GUTHOFF R. F., KOEHLER B., STACHS O.: Characteristic quantities for the quantification of clsm images of the subbasal nerve plexus. *Biomedizinische Technik (Biomedical Engineering)* 55 (2010), 252–254. 2
- [ZPZ*14] ZIEGLER D., PAPANAS N., ZHIVOV A., ALLGEIER S., WINTER K., ZIEGLER I., BRÜGGEMANN J., STROM A., PESCHEL S., KOEHLER B., STACHS O., GUTHOFF R. F., RODEN M.: Early detection of nerve fiber loss by corneal confocal microscopy and skin biopsy in recently diagnosed type 2 diabetes. *Diabetes* (2014), Epub ahead of print. 2
- [ZRD*08] ZIEGLER D., RATHMANN W., DICKHAUS T., MEISINGER C., MIELCK A., KORA STUDY GROUP: Prevalence of polyneuropathy in pre-diabetes and diabetes is associated with abdominal obesity and macroangiopathy – the MONICA/KORA Augsburg surveys S2 and S3. *Diabetes Care* 31, 3 (2008), 464–469. doi:10.2337/dc07-1796. 1
- [ZWH*13] ZHIVOV A., WINTER K., HOVAKIMYAN M., PESCHEL S., HARDER V., SCHÖBER H.-C., KUNDT G., BALTRUSCH S., GUTHOFF R. F., STACHS O.: Imaging and quantification of subbasal nerve plexus in healthy volunteers and diabetic patients with or without retinopathy. *PLoS ONE* 8, 1 (2013), e52157. doi:10.1371/journal.pone.0052157. 2