

Preparing clinical ophthalmic data for research application

Miriam Rößner¹, Stefan Kahl¹, Katrin Engelmann², Danny Kowerko¹

Abstract: This paper presents an analysis of clinical examination, diagnostic and patient data belonging to persons with eye diseases like age-related macular degeneration (AMD). Our purpose is to investigate potential correlations of extracted features to discover their impacts on the disease. This is a first step to the predictability of the progression of AMD based on a heterogeneous data set. We focus on the visual acuity as reasonable indicator for the progression of this disease and analyse its temporal trend to classify patients in winners, stabilisers and losers. We describe the retrieval of textual medical reporting data for optical coherence tomography images and evaluate the machine-readable categorisation of these texts. Additionally, we address the topic of ethical guidelines for the work with patients' data and discuss the potential and limitations of our data set in the context of obtaining structured (mass) data for training neural networks as future perspective.

Keywords: Age-related macular degeneration; Ophthalmology; Text mining; Data visualisation; AMD progression prediction

1 Introduction

In Germany, the rate of blindness increases with age [Wo15]. Until 2013, 20 to 30 % more patients with age-related eye diseases are to be expected. In 2012, 4 884 000 patients in Germany were found to suffer from macular diseases (degenerative and diabetes-induced). Estimations forecast 7.7 millions further patients in the age of 60+. For age-related macular disease (AMD), the cost of ambulant care are about 100 - 200 million Euro per year and another 100 - 150 million Euro for medication. The prevalence (diseased patients) and incidence (newly diseased patients) in Germany are 1.6 million and 305 000 patients, respectively. Common AMD therapies are capable of stabilising or improving, but not of healing the disease. Thus, 5 000 patients per year are expected to become blind. The so-called indirect costs of blindness are not determined in Germany, but estimated to be in a 1:3 ratio to the direct cost [HKN14]. Treatment of patients revealed differences, here coarsely categorised into winner, stabiliser and loser following the nomenclature presented in [Ge11]. There is an increasing demand to categorise patients in objective but also in differentiated manner, what is typically not the case in pharmacological approval studies. A way to tackle this problem is to (digitally) track diagnostic information regularly in

¹ Technische Universität Chemnitz, Junior Professorship Media Computing, Straße der Nationen 62, 09111 Chemnitz, Germany firstname.surname@informatik.tu-chemnitz.de

² Klinikum Chemnitz gGmbH, Augenklinik, Adresse, Germany emailaddress:k.engelmann@skc.de

time facilitating individual therapies that aim to preserve quality of vision for winners and stabilisers and to adapt medication/therapy for losers or low-responders. In our paper, we present a typical workflow to assess and explore clinical data in terms of retrieving machine-readable data allowing for statistical analysis of patient data, e.g. the kind of disease, diagnostic examinations, medication and medical diagnoses/classifications. Secondly, we discuss our current state of analysing more than 30 000 patient data, obtained from our cooperation partner the ophthalmic hospital of the Klinikum Chemnitz gGmbH (OHKC). Results are discussed in the context of recent scientific research projects demonstrating the high potential of convolutional neural networks that were trained with several ten thousands of classified patient data to classify ophthalmic diseases like diabetic retinopathy carried out as competition at Kaggle.com³ and Google Deep Mind's health project [Gu16]⁴. However, it remains unclear whether these techniques can improve current ophthalmologic assessments. Additionally, most research results are inaccessible to the public, which is mainly due to the massive financial potential of automated medical image processing. Google Deep Mind published some results regarding the evaluation of their approach in [Gu16] but did not include any information on algorithms and neural net architectures used. This way, reproducing the results is impossible, emphasising the urgent need for data sets and algorithms available to the research community. Furthermore, processing retinal fundus images is part of a very different domain of medical image data than retinal scans are: Retinal fundus images have higher resolution, sharper contours, significant less noise and, most importantly, are not necessarily suitable for the assessment of AMD progression due to the lack of thickness data.

A technical workflow of transferring and working with clinical data in accordance with ethical guidelines in Germany and the European Union has recently been published elsewhere [Ko17]. To tackle the question of winner, stabiliser and loser classification, we demonstrate limitations and potential of extracting visual acuity (VA) – measurement date time series (VAMDTS) of individual patients. Optical Coherence Tomography (OCT) images are explored in terms of their capability to synchronise them with patient's visual acuity examinations and diagnostic information. We evaluate our developed text-retrieval method in terms of accuracy to extract several AMD-related pathologies on a manually annotated ground truth data test set of 77 examinations from 3 patients.

2 Data set documentation and statistics

The data made accessible by OHKC is exported from two different databases. The first is the OCT device ("Spectralis" by Heidelberg Engineering) internal database that contains all image data of examinations. The device offers the possibility of an anonymised export of E2E files for that it overwrites the patient's name with the corresponding ID. The second database is "TurboMed" by CompuGroup Medical, a medical documentation system that

³ <https://www.kaggle.com/c/diabetic-retinopathy-detection>, 30th of June 2017

⁴ <https://deepmind.com/applied/deepmind-health/>, 30th of June 2017

contains all statements and analysis documented by the physicians. These data can be exported using the standard transfer format bdt. Our challenge is to extract information out of both databases and to combine it to gain knowledge about statistical aspects and correlations of different features.

2.1 Optical coherence tomography data

E2E files contain all examination data of one patient collected by the OCT device (in OHKC). One OCT scan produces 25 slices per examination and a fundus image including markings for the slice's positions. Both image types are readable and exportable using the Unified OCT explorer (UOCT) [Ro16]. Each slice exists as raw data and preprocessed data. Additionally two contours per slice are given. They mark the retinal layers “Bruch Membrane” (BM) and “Inner Limiting Membrane” (ILM) that enclose the retina and can be used for determining its thickness. The following metadata are included: the 8-digit hospital intern patient ID, 4-digit device internal patient ID, birth date, sex, laterality of each scan (left or right eye) and the corresponding examination dates. The contours are generated automatically by the device and can be corrected manually by the physicians. Because false contours are not always corrected the specifications are not 100 % reliable. Figure 1 shows an extracted OCT scan and corresponding contours. Approaches of improving BM and/or ILM localisation using image processing have been reported elsewhere [KRR14, Hu13].

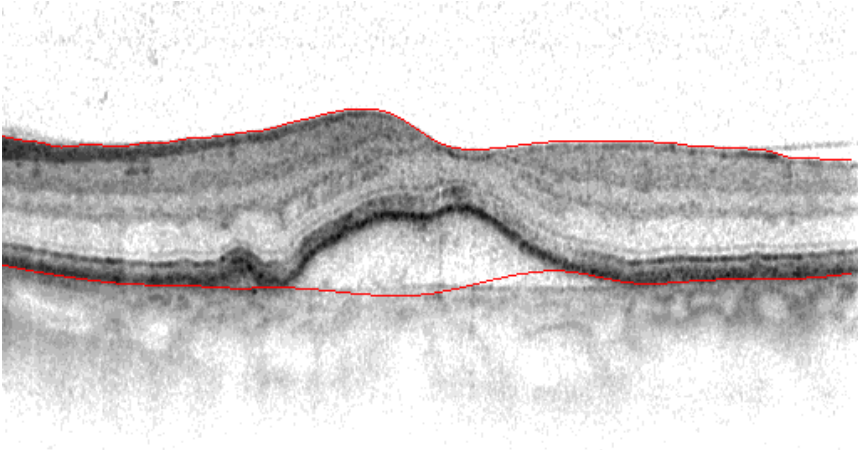


Fig. 1: Central OCT scan of a left eye with corresponding contours (red, ILM above, BM below)

2.2 Patient and diagnostic data

TurboMed is the medical documentation system used in the OHKC. It contains a collection of all examination and treatment data of every patient. The system offers the possibility to export the data using the transfer format bdt. The resulting export is a labelled text file. Every line contains a numeric code that specifies the line's content. For details see the official format description [IT16]. TurboMed as well as the bdt export file are divided in different sections containing different types of data, e.g. tables for visual acuity or listings of intended treatment appointments. The section we focus on in the first place is the section for text entries of different types defined by hospital internal standards. In bdt format the type of this section is indicated by the numerics 6330 followed by the specified section heading. The following lines labelled with 6331 contain the entry's content as continuous text. We are particularly interested in the entries labelled as OCT statements (heading: OCT). As important information for categorisation, we also extract the visual acuity tables. This data shows a more objective way to describe the course of disease (clear numbers) than the OCT statements that are more difficult to classify (text format). To make the text statements classifiable we convert them to fit into our developed uniform scheme shown in table 1.

Tab. 1: Classification categories for OCT reports

Category	Possibilities
Intraretinal fluid	true, false
Subretinal fluid	true, false
Subretinal fibrosis	true, false
RPE-detachment	true, false
Foveal depression	true, false
Central retina thickness	extracted value
Cicatrice	true, false
RPE layer	continuous, discontinuous
ELM layer	continuous, discontinuous
Ellipsoid layer	continuous, discontinuous

2.3 Fusion of OCT and diagnostic data

The following medical report belongs to the scan in Figure 1:

“R/L persistierende RPE-Abhebungen zentral, kein Anhalt für sub- oder intraretinale Flüssigkeit, zentrale Netzhautdicke R 330 μm L 354 μm ”

Translated this statement means:

“Right/Left persistent central RPE-detachment, no evidence of sub- or intraretinal fluid, central retina thickness right 330 μm left 354 μm ”

After processing using text mining, different categories were determined:

RPE-detachment: true

Subretinal fluid: false

Intraretinal fluid: false

Central retina thickness: 354

The correlation between OCT scan and medical report was found because of the patient ID and the examination date. In the E2E file the date is given as the Unix timestamp 1377509482668, which is the date 26th of August 2013. This date could be compared with all dates of examination reports for the corresponding patient ID to find the match.

We can also state that the date of about 70 % of the extracted OCT reports matches exactly with the date of a visual acuity measurement. This was determined by analysing the data of the 10 patients with the most OCT reports and their corresponding extracted visual acuity data. These are only the exact matches, but there are several more entries with a date that is one day before or after the report.

2.4 Statistics of the OHKC data set

More than 31 300 patient data from the OHKC build the base of our data set. A histogram of all patients' date of birth is shown in the left of Figure 2 together with the total number of patients sorted by gender that have been treated since 2002. Obviously, the data reveal gender equity and of course an increasing fraction of patients towards the old age.

The cumulative number of patients as function of VA measurement period is illustrated in the right of Figure 2, where about half of the patients appear only with 1 VA measurement. Patients with visual acuity courses longer than 1 year are marked by dotted lines. For patients with visual acuity (VA) courses longer than 1 year and at least 5 measurements, the VA courses for 9 patients from 3 groups are exemplarily illustrated in Figure 3 (winner, stabiliser, loser: from top to bottom), where each data set is approximated by a linear equation to get a coarse indication whether the VA improves, remains constant or decreases. The mean squared error (MSE) of each linear regression provides information about the accuracy of the underlying model. In case of large MSE values equations different from being linear might be more plausible. As the VA is rather expected to have an upper and lower boundary, sigmoidal-type of functions might describe some VA courses better, e.g. the ones in the top right and bottom right of Figure 3.

The slopes derived from linear regression in Figure 3 was converted into the more intuitive unit of VA change per year. Its cumulative frequency is illustrated in Figure 4 for all patients with VA courses longer than 1 year and at least 5 measurements. The diagram reveals that a fraction of 30–40 % of these roughly 2 000 patients exhibits a loss of visual acuity. Note,

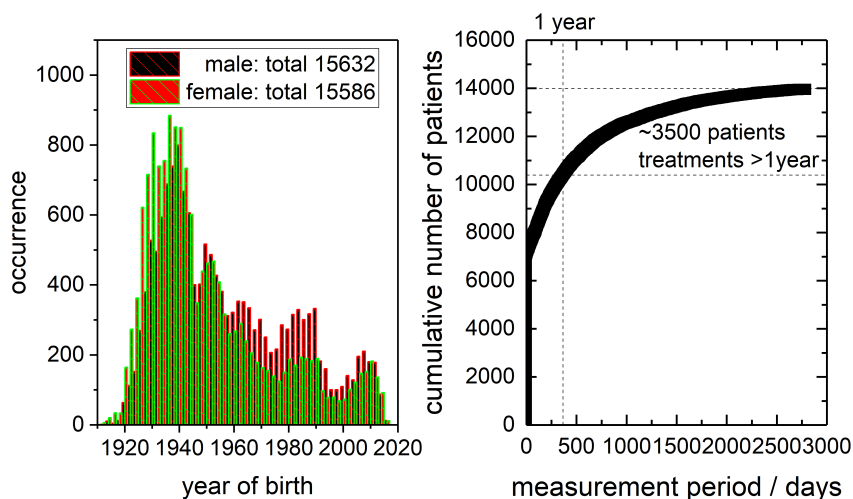


Fig. 2: Statistical data of the OHKC data set: (left) Distribution of year of birth from all patients treated or examined during the years 2002 to 2017 and (right) Cumulative distribution of patients with increasing examination time, here defined by the period between initial and last measurement of visual acuity.

that to this end, we have no systematic data on medication or whether there is a therapy in progress or not.

3 Evaluation of AMD-relevant bdt-export fields

The evaluation of the patient data extraction shows a big difference based on the data type. The fields birth date and sex are very well machine-readable, because they are short values that do not vary in form. This is proven by the evaluation showing that 100 % of the first 20 patients (sorted by IDs) got the data attached correctly. The visual acuity is nearly as well machine-readable as the personal data. It is listed in tables that easily can be read out. Unfortunately many empty tables are existent because the visual acuity values are not continuously documented. Hence we could only extract visual acuity data of about 15 000 patients.

The evaluation of the extraction of OCT reports reveals more difficulties. For the evaluation, the 10 biggest extracted files (10 to 16 reports each, one file per patient) were compared to the original data. We were able to extract nearly 80 % of all reports. The rest is difficult to extract because there is no separate field for the reports. It is a common field that can be labelled as OCT report via a heading. But that heading differs and some OCT reports are not even marked as such. In total we are able to extract OCT reports of about 5 000 patients.

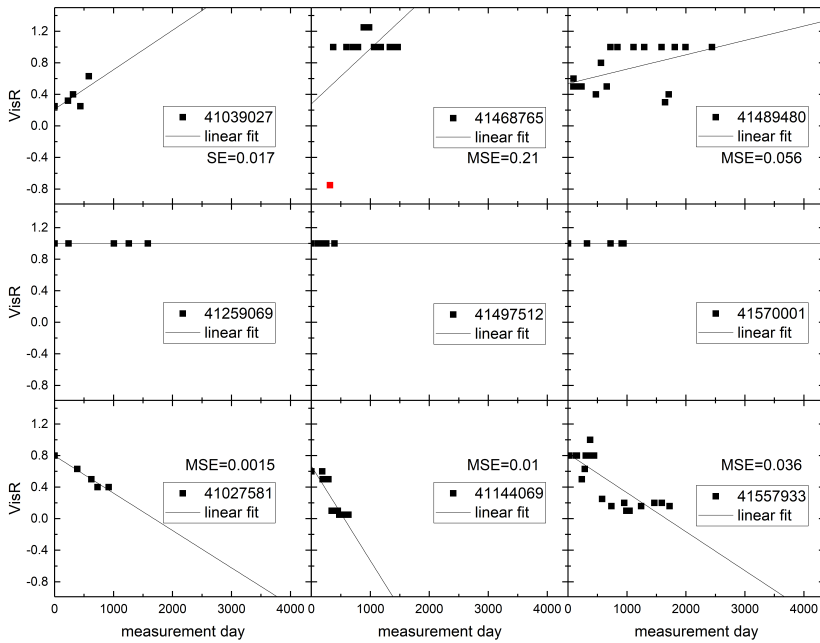


Fig. 3: Examples of 9 patients' individual progression of VA with 3 categories: deterioration, stabilisation and improvement (bottom to top).

It is likely that all other patients got a different diagnosis than AMD. We also can not say definitely that all of those 5 000 patients got AMD. To classify the report data we used text retrieval to cope with the complexity of the text. For the evaluation of the classification, data of 3 patients with 77 reports in total (38 reports containing data for right and left eye + one report containing data for only one eye) were analysed and corrected to generate a ground truth. For the analysis the occurrence of all categories in the extracted file was compared with the occurrence in the result files to determine the number of false negatives. Also the correctness of the classification was tested to determine the false positives. The complete process took about four hours of work for one person. The results are shown in table 2.

It can be seen that the most frequent categories are central retina thickness, sub- or intraretinal fluid, foveal depression and RPE-detachment. Those categories show a high reliability in classification. That means that the used method can be applied to the total data set to generate a bigger ground truth for these categories (with about 90 % reliability instead of 100 %). The problematic categories about the continuity of the layers should not be used as ground truth before the used methods are improved. Those categories are less frequent

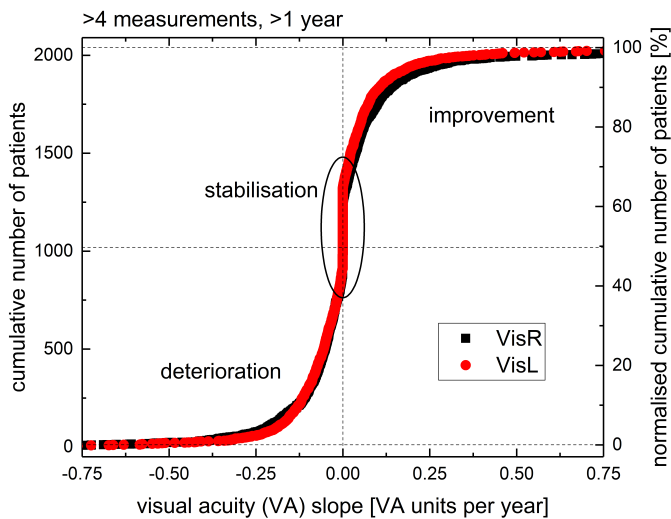


Fig. 4: Cumulative distribution VA slopes of patients with at 5 VA measurements spread over at least one year.

Tab. 2: Evaluation results of the classification of OCT reports

Category	Ground truth	False positive	False negative
Intraretinal fluid	57	2 (3.5 %)	7 (12.3 %)
Subretinal fluid	48	1 (2.1 %)	0
Subretinal fibrosis	5	0	0
RPE-detachment	23	0	1 (4.3 %)
Foveal depression	47	0	3 (6.4 %)
Central retina thickness	59	4 (6.8 %)	0
Cicatrice	9	0	0
RPE layer	3	0	3 (100.0 %)
ELM layer	3	0	2 (66.7 %)
Ellipsoid layer	5	0	4 (80.0 %)

because they are new to the documentation. Likewise, there are varying expressions used that makes a text retrieval difficult.

4 Conclusion and Outlook

We presented a workflow of retrieving and processing medical raw data to make them machine-readable for research purposes. As our approach relies on common data formats like *.e2e from the wide-spread Heidelberg OCT device and *.bdt, a standardised and documented exchange format for patient data including diagnostic information, we assume our strategy to be scalable and applicable to other ophthalmic hospitals. From more than 30 000 formally available patient data, typically only subsets are usable for research due to different documentation status, e.g. of visual acuity or due to individual textual descriptions of medically equivalent diagnoses of pathologies. Within the OHKC data set more than 2 000 VAMDTs (visual acuity courses) longer than one year and 5 measurements revealed that about 60–70 % of the patients from this data set may be classified as stabiliser or winners during the course of therapy. However, about 30 % of patients with decreasing visual acuity over time were identified as subject for deeper inspection in the future, looking for the answer why their response is low.

From the diagnostic bdt-fields, a unique database scheme was derived and a test set of 77 OCT reports manually adapted to define the ground truth for the evaluation of our Python-based medical text retrieval method. Reasonably good results were obtained for AMD-relevant classifications like sub- or intraretinal fluid and central retina thickness, while other categories like the layer descriptions are subject to misinterpretations and require manually-assisted or more advanced text retrieval techniques, like recently demonstrated by the winners of task 1 of the CLEF eHealth Evaluation 2016 where written free-text records had to be structured automatically into a given scheme by identifying relevant text-snippets [Eb16]. The selected 2 000 patients VAMDTs inherently comprise more than 10 000 OCT volume-scans giving a reasonable number of training data for neural networks, that could assist automatic classification of pathologies giving recommendations or spatially marked regions important for early identification of the low-responder group.

Acknowledgement

We thank Dr. Olga Furashova and the whole team of the OHKC for their kind support. This work was partially accomplished within the project *localizeIT* (funding code 03IPT608X) funded by the *Federal Ministry of Education and Research* (BMBF, Germany) in the program of *Entrepreneurial Regions InnoProfile-Transfer*. The European Union and the European Social Fund for Germany partially funded this research.

References

- [Eb16] Ebersbach, Mike; Herms, Robert; Lohr, Christina; Eibl, Maximilian: Wrappers for Feature Subset Selection in CRF-based Clinical Information Extraction. In: CLEF (Working Notes). pp. 69–80, 2016.

- [Ge11] Gerding, Heinrich; Loukopoulos, Vlassios; Riese, Juliane; Hefner, Lars; Timmermann, Melanie: Results of flexible ranibizumab treatment in age-related macular degeneration and search for parameters with impact on outcome. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 249(5):653–662, May 2011.
- [Gu16] Gulshan, Varun; Peng, Lily; Coram, Marc; Stumpe, Martin C; Wu, Derek; Narayanaswamy, Arunachalam; Venugopalan, Subhashini; Widner, Kasumi; Madams, Tom; Cuadros, Jorge et al.: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22):2402–2410, 2016.
- [HKN14] Hirneiß, C.; Kampik, A.; Neubauer, A.S.: Volkswirtschaftliche Kosten von Augenerkrankungen. *Der Ophthalmologe*, 111(5):420–427, May 2014.
- [Hu13] Huang, Yijun; Danis, Ronald P.; Pak, Jeong W.; Luo, Shiyu; White, James; Zhang, Xian; Narkar, Ashwini; Domalpally, Amitha: Development of a Semi-Automatic Segmentation Method for Retinal OCT Images Tested in Patients with Diabetic Macular Edema. *PLoS ONE*, 8(12):e82922, December 2013.
- [IT16] IT in der Arztpraxis: Datensatzbeschreibung KVDT, 2016.
- [Ko17] Kowerko, Danny; Rößner, Miriam; Kahl, Stefan; Herms, Robert; Eibl, Maximilian; Engelmann, Katrin: Aufbereitung augenmedizinischer Bild-, Patienten- und Diagnosedaten zum Zwecke der Forschung - Ethikrichtlinien und deren praktische Umsetzung. In: *Mensch und Computer Workshop: Ethische Herausforderungen in sozioinformatischen Forschungsprojekten*. Regensburg, p. submitted, September 2017.
- [KRR14] Kahl, Stefan; Ritter, Marc; Rosenthal, Paul: Automatisierte Beurteilung der Schädigungssituation bei Patienten mit altersbedingter Makuladegeneration (AMD). In (Publishing, KIT Scientific, ed.): *Proceedings of Forum Bildverarbeitung*. Karlsruhe, pp. 179–190, November 2014.
- [Ro16] Rosenthal, Paul; Ritter, Marc; Kowerko, Danny; Heine, Christian: OphthalVis - Making Data Analytics of Optical Coherence Tomography Reproducible. In (Lawonn, Kai; Hlawitschka, Mario; Rosenthal, Paul, eds): *EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization (EuroRV3)*. The Eurographics Association, Groningen, October 2016.
- [Wo15] Wolfram, C: Demografie und Altersabhängigkeit bei Augenerkrankungen. *Der Ophthalmologe*, 1(112):73–81, 2015.