

# Clinical Decision Support for the Classification of Diabetic Retinopathy: A Comparison of Manual and Automated Results

Christoph MITSCH<sup>a,1</sup>, Karsten FEHRE<sup>b</sup>, Sonja PRAGER<sup>a</sup>, Christoph SCHOLDA<sup>a</sup>,  
Katharina KRIECHBAUM<sup>a</sup>, Thomas WRBA<sup>c</sup> and Ursula SCHMIDT-ERFURTH<sup>a</sup>

<sup>a</sup>Department of Ophthalmology, Medical University of Vienna

<sup>b</sup>Medexter Healthcare GmbH, Vienna

<sup>c</sup>Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna

**Abstract.** The management of diabetic retinopathy, a frequent ophthalmological manifestation of diabetes mellitus, consists of regular examinations and a standardized, manual classification of disease severity, which is used to recommend re-examination intervals. To evaluate the feasibility and safety of implementing automated, guideline-based diabetic retinopathy (DR) grading into clinical routine by applying established clinical decision support (CDS) technology. We compared manual with automated classification that was generated using medical documentation and an Arden server with a specific medical logic module. Of 7169 included eyes, 47% (n=3373) showed inter-method classification agreement, specifically 29.4% in mild DR, 38.3% in moderate DR, 27.6% in severe DR, and 65.7% in proliferative DR. We demonstrate that the implementation of a CDS system for automated disease severity classification in diabetic retinopathy is feasible but also that, due to the highly individual nature of medical documentation, certain important criteria for the used electronic health record system need to be met in order to achieve reliable results.

**Keywords.** Diabetes complications, retina, decision support systems, clinical.

## 1. Introduction

Diabetic retinopathy (DR) is the leading cause of visual impairment in working-age adults worldwide [1, 2]. The disease is characterized by capillary non-perfusion and ischemia within the retina, which ultimately leads to macular edema and retinal neovascularization with the potential to severely damage visual function [3]. Although the development of sight-threatening complications of diabetes mellitus (DM) can be delayed by appropriate treatment of systemic diseases such as DM itself, high blood pressure and lipid metabolism abnormalities [4], an estimated 40% (8% for vision-threatening DR) of people with type 2 diabetes and 86% (42%) with type 1 diabetes in the USA have diabetic retinopathy [5, 6].

With the proportion of people who have a predominantly sedentary lifestyle and are overweight growing worldwide, the prevalence of type 2 DM and thus of DR is

---

<sup>1</sup> Corresponding Author: Dr. C. Mitsch, Department of Ophthalmology, Medical University of Vienna, Waehringer Guertel 18-20, 1090 Vienna. E-Mail: christoph.mitsch@meduniwien.ac.at

continuously increasing, burdening healthcare systems all over the world. Patients with DM and DR need to be examined regularly by various specialists: At least one general practitioner and one ophthalmologist are involved in a patient's routine check-ups. In the case of ophthalmological examinations, indirect ophthalmoscopy (examination of the patient's retinae through dilated pupils) is the only way to determine the extent of retinopathy. This examination is time-costly and requires a specialist; it is not possible to delegate it to non-medical personnel.

According to the American Academy of Ophthalmology guidelines for the management of DR [7], follow-up intervals should depend on the current stage of disease progression. Patients with normal retinae or mild DR should be re-examined annually, while those with moderate DR should be re-examined within 6 to 12 months as disease progression is common. Patients with severe DR should be re-examined within 2-4 months and patients with proliferative DR should receive treatment with appropriate follow-up intervals.

Medical informatics in general and clinical decision support systems in particular aim to assist healthcare delivery and are especially pertinent when a high patient load meets with well-defined clinical processes. With the growing incidence of DM, it seems logical to make use of computer technologies. Many efforts have been reported: Automated, fundus-photograph-based disease detection and classification (also termed automated retinal image analysis, ARIA; or computer-aided detection; CAD) is a promising technique to relieve healthcare systems' burdens of regular screening of diabetic patients. Many software products offering this technology are commercially available and could potentially improve the manner in which diabetes eye care is delivered [8]. Automated detection of more advanced diabetic lesions in the patient's eye has become conceivable with the advent of new, non-invasive optical coherence angiography. Integration of the resulting data with information about the patients' systemic disease such as diabetes duration, hemoglobin A1c, and body mass index, and demographics such as age and sex allows predictive modeling to be applied and thus prognosis of the individual disease course [9].

The Arden syntax is a standard for representing clinical and scientific knowledge in an executable format. The Arden server stores and processes details from medical guidelines in the form of so-called medical logic modules (MLMs), and provides an interface for administrative and maintenance tasks [10]. The server is used in a wide spectrum of clinical situations such as adverse drug event monitoring [11] or medical guideline implementation [12]. Although this technology is not new, there are few clinical decision support applications in the field of ophthalmology and most of them are related to automated image analysis or guides and material for preparing ophthalmology students [13].

The implementation of technologies that potentially increase efficiency in the management of chronic diseases is highly attractive, especially when it enables maintaining consistent quality of medical care.

We are not aware of any published efforts to use clinical decision support (CDS) technology in the classification of DR. Thus, the aim of this study was to compare the results of manual classification of DR with those of a CDS-based automated classification and assess the applicability and safety of the CDS-based automated with the manual method for disease classification before deploying this technology for its use in daily clinical routine at our department.

## 2. Methods

### 2.1. Classification of DR

To date, the International Clinical Disease Severity Scale for DR[14] has become the most commonly used severity classification of DR. The variables used are related to the integrity of and pathological alterations to retinal vessels which specialists observe on the retinae of the patients via ophthalmoscopy. Their presence and extension are noted and manually processed as described in Table 1. This manual approach currently represents the gold standard of DR classification.

**Table 1** - The International Clinical Disease Severity Scale for diabetic retinopathy (DR). IRMA = intraretinal microvascular abnormalities

| <b>Disease Severity Level</b> | <b>Findings Observable upon Dilated Ophthalmoscopy</b>  |
|-------------------------------|---|
| No retinopathy                | No abnormalities  |
| Mild DR                       | Microaneurysms only   |
| Moderate DR                   | More than just microaneurysms but less than severe DR   |
| Severe DR                     | Any of the following: <ul style="list-style-type: none"> <li>• More than 20 intraretinal hemorrhages in each of four quadrants</li> <li>• Venous beading in two or more quadrants</li> <li>• Prominent IRMA in one or more quadrants</li> </ul> .. and no signs of proliferative DR |
| Proliferative DR              | One or both of the following: <ul style="list-style-type: none"> <li>• Neovascularization</li> <li>• Vitreous/pre-retinal hemorrhage</li> </ul>   |

### 2.2. Ophthalmological findings

*Microaneurysms* (MA) are small alterations to retinal capillaries which, clinically, mark the very beginning of retinal disease in patients with DM. The rupture of MA leads to *dot-blot hemorrhages*. Areas of venular dilatation and bulging are called *venous beading* and *Intra-retinal microvascular abnormalities (IRMA)* are newly grown vessels within the layers of the retina. Extra-retinal neovascularization can occur at the optic disc (*neovascularization on the optic disc*, NVD), elsewhere on the retina (*neovascularization elsewhere*, NVE) or in the anterior chamber of the eye (*neovascularization on the iris*, NVI). Rupture of newly formed vessels can result in *pre-retinal or vitreous hemorrhages*.

### 2.3. Documentation of relevant findings

Clinical examinations and observations at the outpatient clinic for Diabetic Retinopathy and Traumatology of the Department of Ophthalmology and Optometry of the Medical University of Vienna are performed and documented either by resident physicians working under consulting physicians' supervision or by experienced consulting physicians themselves. Information is stored in a customized electronic health record (EHR) system. This system consists of a series of forms on the Research Documentation and Analysis platform (RDA) of the Medical University of Vienna's center for medical

**Table 2** - List of the variables relevant for the classification of diabetic retinopathy. P = present, P1 = present in one quadrant, P2+ = present in more than one quadrant, A = active, F = fibrotic, FT = fibrotic with traction, NP = not present, NA = not assessed, IMP = impossible to assess

| Observation                              | Values  |
|--|---|
| Microaneurysms                           | P, NP, NA, IMP  |
| Rubeosis                                 | P, NP, NA, IMP  |
| Venous tortuosity                        | P1, P2+, NP, NA, IMP  |
| Neovascularization at the optic disc     | P, A, F, FT, NP, NA, IMP  |
| Neovascularization elsewhere             | P, A, F, FT, NP, NA, IMP  |
| Intraretinal microvascular abnormalities | P, NP, NA, IMP  |
| Retinal hemorrhages                      | Present less than 20x in all quadrants, present at least 20x per quadrant, P, NP, NA, IMP |
| Vitreous or pre-retinal hemorrhage       | P, NP, NA, IMP  |

statistics, informatics and intelligent systems (CEMSIIS). Besides general functional and morphologic data, findings relevant for diabetic retinopathy are stored in a highly structured manner, separately for each eye. See Table 2 for a list of the variables queried.

#### 2.4. Manual documentation of diabetic retinopathy severity

Retina specialists manually documented a retinopathy classification separately for each eye of patients in case the severity of the present diabetic retinopathy was relevant. It is important to note that our EHR does not require a disease classification, thus, it is not possible to differentiate between no present retinopathy and retinopathy not assessed. Two further possible values of the manual classification were “not graded/no retinopathy”, “retinopathy; severity not classified”. These values were not compared with automatic results. The descriptors for the different stages of diabetic retinopathy were “mild”, “moderate”, “severe”, and “proliferative” RP, the last stage further described as “active proliferative” or “inactive proliferative”, or “fibrotic proliferative RP”.

**Table 3** - The algorithmic part of the Arden medical logic module used for automated diabetic retinopathy severity classification. For demonstration purposes, the header and the code for the left eye have been removed and the values translated into the English language. The suffix “\_od” indicates that the variable represents a finding on the right eye (o.d. = *oculus dexter*)

| Diabetic Retinopathy Medical Logic Module (right eye)  |
|--|
| <pre> IF ((vitreous_hem_od = "true") OR (nvd_od = "true") OR (nvd_od = "active") OR (nvd_od="fibrotic") OR (nvd_od = "fibrotic with traction") OR (nve_od = "true") OR (nve_od = "active") OR (nve_od="fibrotic") OR (nve_od = "fibrotic with traction") OR (rub_od = "true")) THEN     result.ScoreOD := 4; # proliferative DR ELSEIF ((venous_tortuosity_od = "Present in more than one quadrant") or (hemorrhage_od = "true, present at least 20x per quadrant") or (irma_od = "true")) THEN     result.ScoreOD := 3; # severe DR ELSEIF (hemorrhage_od = "true, present less than 20x in all quadrants") THEN     result.ScoreOD := 2; # moderate DR ELSEIF (ma_od = "true") THEN     result.ScoreOD := 1; # mild DR ELSEIF (ma_od = "false") THEN     result.ScoreOD := 0; # no DR ELSE     result.ScoreOD := -1; # not enough data for automated grading ENDIF; </pre> |

## 2.5. Automated classification of diabetic retinopathy severity

See Table 3 for the medical logic module (MLM) programmed for the automated classification. The possible results of the algorithm were “not enough data for automated grading”, “no retinopathy”, “mild DR”, “moderate DR”, “severe DR” and “proliferative DR”.

## 2.6. Inclusion and exclusion criteria

Visit documentations which included patho-morphological findings and a manual DR grading were included. We excluded entries which had either no grading or no findings documented.

## 3. Results

### 3.1. Overview

We analyzed the documentations of 5727 visits of 1303 out-patients at the department of diabetic retinopathy and traumatology between 2012 and 2015, relating to both eyes, resulting in a dataset of 11454 eye examinations. Explicit manual classification (a diabetic retinopathy level of at least “mild”) had been performed for 7530 eyes. Based on the documented morphology findings, automated classification was possible in 10293 eyes and not possible in 1161 eyes. No manual classification had been performed for 3924 eyes. This left us with an includable subset of 7169 eyes, where manual and automated classification were available and possible.

### 3.2. Included visits: Manual classification performed, automated classification possible

Based on the manual classification, 14.2% (n=1020) showed mild DR, 23.2% (n=1666) moderate DR, 18.6% (n=1335) severe DR, and 43.9% (n=3148) proliferative DR. The results of the automated classification were 17.8% (n=1277) with no DR, 15.6% (n=1117) with mild DR, 20.2% (n=1450) with moderate DR, 10.6% (n=762) with severe DR and 35.8% (n=2563) with proliferative DR. See the left graph on Figure 1 for a graphical representation of those distributions. The overall inter-method agreement was 47%; 29.4% in mild DR, 38.3% in moderate DR, 27.6% in severe DR, and 65.7% in proliferative DR. See the right graph on Figure 1 for a graphical representation of the agreement levels. The numbers and percentages (relative to the included cohort) of the eyes with equal and differing classifications are given in Table 4. The greatest fraction with disagreement was one group of eyes which had been manually graded as showing signs of “proliferative DR”, but calculated as “no DR” disease by the automatic algorithm, constituting 6.4% (n=459) of all eyes.

The most frequent relative comparison result was by far “agreement” (difference = 0 stages, 47%, n=3373). 18.3% (n=1080) of all eyes compared were classified with a more severe DR automatically than manually, and 37.9% (n=2716) with a less severe DR. Classified automatically, this manual “less severely classified” group consisted of 5.2% (n=376) with moderate DR, 2.9% (n=208) with severe DR and 6.9% (n=496) with proliferative DR. Refer to the middle graph on Figure 1 for a graphical representation of

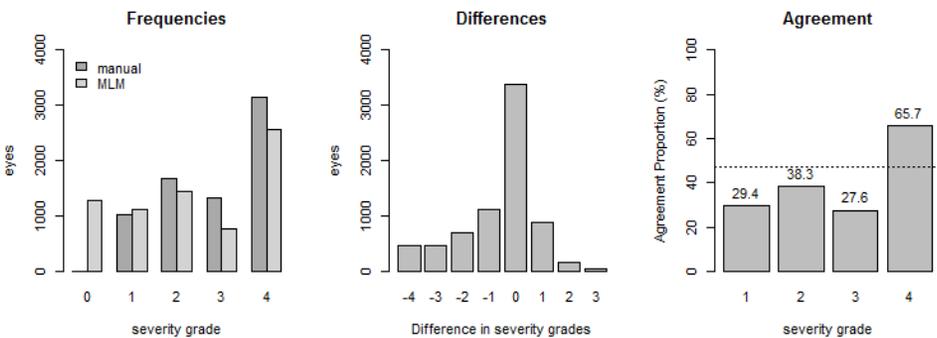
**Table 4** – Proportions (relative to the whole cohort) of the groups with agreement (light grey cell background) and disagreement (white cell background) of all manual and all automated classification results. Rows: Manual classification M1 (mild DR) to M4 (proliferative DR). Columns: Automated classification A0 (no DR) to A4 (proliferative DR).

|    | A0           | A1           | A2           | A3           | A4             |
|----|--------------|--------------|--------------|--------------|----------------|
| M1 | 3.8% (n=275) | 4.2% (n=300) | 5.2% (n=376) | 0.4% (n=32)  | 0.5% (n=37)    |
| M2 | 4.3% (n=306) | 5.9% (n=421) | 8.9% (n=638) | 2.5% (n=176) | 1.7% (n=125)   |
| M3 | 3.3% (n=237) | 2.4% (n=172) | 3.1% (n=224) | 5.1% (n=368) | 4.7% (n=334)   |
| M4 | 6.4% (n=459) | 3.1% (n=224) | 3% (n=212)   | 2.6% (n=186) | 28.8% (n=2067) |

the frequencies of the relative classification differences. An analysis of the categorization differences between the individual consulting physicians revealed the same range (-4 to 3 degrees) and a mean grading difference of  $-0.62 \pm 1.42$  degrees for all graders.

3.3. Eyes excluded due to missing manual grading or unfeasible automated grading

No manual classification information was available for 34.3% (n=3924) of all screened eyes. This group comprised all eyes which had not presented with DR or no classification had been documented. For 27.3% (n=3124) of the eyes, no manual grading had been performed, but automated classification was possible. No findings relevant to DR had been documented, thus no DR was automatically classified for 18.5% (n=2123), mild DR for 2.1% (n=235), moderate DR for 2.6% (n=300), severe DR for 1.1% (n=127) and proliferative DR for 3% (n=339) of all eyes. Automated classification was not possible in 10.1% (n=1161) of all eyes due to missing findings documentation. In 3.2% (n=361) of all eyes automated classification was not possible, but manual classification had been performed. In this group, expressed as proportions of all screened eyes, 0.6% (n=66) had mild RP, 0.7% (n=77) moderate RP, 0.9% (n=102) severe RP and 1% (n=116) proliferative disease. No manual or automated classification was available for 7% (n=800) of all screened eyes.



**Figure 1** – Results of the manual and automated classifications (left graph), a plot of the differences and their frequencies (middle graph) and a bar-plot of the agreements in the different severity classes. Left: 0 = no DR, 1 = mild DR, 2 = moderate DR, 3 = severe DR, 4 = proliferative DR. Middle: x-Axis = classification difference between automated and manual classification. Positive values suggest a worse automated disease severity, negative values a lesser automated severity. Right: 1 = mild DR, 2 = moderate DR, 3 = severe DR, 4 = proliferative DR. The dotted horizontal line represents the overall agreement.

#### 4. Discussion

We compared the manual and CDS-based automated DR classification of 7169 eyes examined at our diabetic retinopathy and traumatology outpatient clinic. The overall agreement between the methods was 47%, which, essentially, is not very satisfactory, given that reevaluation interval recommendations are potentially determined based on their disease classification. We used data generated by ophthalmologists in our diabetic retinopathy outpatient clinic, which, being a tertiary care center, mainly focuses on the management of more advanced disease. This may reflect in the inter-method agreement, which is the highest (65.7%) amongst proliferative cases. Retreatment intervals at this stage are much shorter and set individually, based on planned and performed interventions. Our custom EHR does not validate or check the fields required for disease staging for completeness. Morphological findings, disease classification and reexamination interval represent three independent, but, from a physician's perspective, rather redundant data elements. They are inherently linked by guideline recommendations. The high patient load in this kind of outpatient clinics does not always allow to document redundantly, so the most important piece of information remains whether an intervention is necessary or not, and, if no intervention is indicated, when an ophthalmological reexamination needs to take place. In our custom EHR, this is written as a free-text message to collaborating field ophthalmologists. Thus, the structured analysis of this information was not easily possible.

The inter-method agreement presented in this experiment has wide potential of improvement. In order to achieve sufficient reliability of automated disease classification, the input data and therefore our EHR needs to meet specific criteria. The lessons we learned are the following: First, and most importantly, redundancy needs to be eliminated. This is a general principle, not only for medical settings, and has a very specific meaning in the EHR context: If CDS is to be implemented in clinical routine and to be working with data documented during patient visits, the forms need to be reduced to only offer the one most precise specification of the relevant parameters (those in Table 2). If the classification is straight-forward and free of exceptions, which is the case in this example, the input of manual results should be similarly possible, but linked to an informative validation of its consistency with the input parameters (the documented morphological findings). The re-examination interval recommendation should be offered as a suggestion which can be altered later, as there may be other factors taken into account for this decision. If manually entered data is used for the automated calculation of patient management recommendations, it needs to be automatically validated for consistency and completeness before CDS gets to work, which, in our case, would mean checking all relevant entered ophthalmological findings data.

The main challenge seems to be preserving EHR usage flexibility for the physicians. In contrast to medical guidelines, medical documentation is a highly subjective and personal task. CDS should be implemented in a facultative manner, so that physicians can decide whether they want to use it (in which case they would have to follow certain documentation rules, see above), or if they just want to document the most relevant findings, determine the disease classification and reevaluation intervals by themselves, which is what we see in our present data.

We believe that, if those rules are met, CDS-based disease classification is certainly possible for clinical routine, where it can contribute to uniform and objective assessment and help to save valuable physicians' time, which can then be spent with the patient rather than with a computer.

To conclude, we demonstrate that the implementation of a CDS system for automated disease severity classification in an outpatient clinic treating patients with diabetic retinopathy is technically and medically feasible but that specific adaptation to the EHR structure is required in order to yield reliable results.

## References

- [1] J. E. Shaw, R. A. Sicree, and P. Z. Zimmet, "Global estimates of the prevalence of diabetes for 2010 and 2030," *Diabetes Res. Clin. Pract.*, vol. 87, no. 1, pp. 4–14, 2010.
- [2] R. Klein, K. E. Lee, M. D. Knudtson, R. E. Gangnon, and B. E. K. Klein, "Changes in visual impairment prevalence by period of diagnosis of diabetes: the Wisconsin Epidemiologic Study of Diabetic Retinopathy.," *Ophthalmology*, vol. 116, no. 10, pp. 1937–42, 2009.
- [3] D. A. Antonetti, R. Klein, and T. W. Gardner, "Diabetic retinopathy.," *N. Engl. J. Med.*, vol. 366, no. 13, pp. 1227–39, Mar. 2012.
- [4] D. Wright and P. M. Dodson, "Medical management of diabetic retinopathy: fenofibrate and ACCORD Eye studies.," *Eye (Lond.)*, vol. 25, no. 7, pp. 843–849, 2011.
- [5] J. H. Kempen, B. J. O'Colmain, M. C. Leske, S. M. Haffner, R. Klein, S. E. Moss, H. R. Taylor, and R. F. Hamman, "The prevalence of diabetic retinopathy among adults in the United States," *Arch. Ophthalmol.*, vol. 122, no. 4, pp. 552–563, 2004.
- [6] M. S. Roy, R. Klein, B. J. O'Colmain, B. E. K. Klein, S. E. Moss, and J. H. Kempen, "The prevalence of diabetic retinopathy among adult type 1 diabetic persons in the United States.," *Arch. Ophthalmol. (Chicago, Ill. 1960)*, vol. 122, no. 4, pp. 546–51, Apr. 2004.
- [7] American Academy of Ophthalmology Retina/Vitreous Panel, "Preferred Practice Pattern (R) Guidelines. Diabetic Retinopathy. San Francisco, CA: American Academy of Ophthalmology; 2014. Available at: [www.aao.org/ppp](http://www.aao.org/ppp)."
- [8] D. A. Sim, P. A. Keane, A. Tufail, C. A. Egan, L. P. Aiello, and P. S. Silva, "Automated retinal image analysis for diabetic retinopathy in telemedicine.," *Curr. Diab. Rep.*, vol. 15, no. 3, p. 14, 2015.
- [9] T. Aspelund, O. Thornórisdóttir, E. Olafsdóttir, a Gudmundsdóttir, a B. Einarssdóttir, J. Mehlsen, S. Einarsson, O. Pálsson, G. Einarsson, T. Bek, and E. Stefánsson, "Individual risk assessment and information technology to optimise screening frequency for diabetic retinopathy.," *Diabetologia*, vol. 54, no. 10, pp. 2525–32, 2011.
- [10] M. Samwald, K. Fehre, J. de Bruin, and K. P. Adlassnig, "The Arden Syntax standard for clinical decision support: Experiences and directions," *J. Biomed. Inform.*, vol. 45, no. 4, pp. 711–718, 2012.
- [11] K. Fehre, M. Plössnig, J. Schuler, C. Hofer-Dückelmann, A. Rappelsberger, and K. P. Adlassnig, "Detecting, Monitoring, and Reporting Possible Adverse Drug Events Using an Arden-Syntax-based Rule Engine," *Stud. Health Technol. Inform.*, vol. 216, no. 4, p. 950, 2015.
- [12] A. Seitingner, K. Fehre, K.-P. Adlassnig, A. Rappelsberger, E. Wurm, E. Aberer, and M. Binder, "An Arden-Syntax-based clinical decision support framework for medical guidelines--Lyme borreliosis as an example.," *Stud. Health Technol. Inform.*, vol. 198, pp. 125–32, 2014.
- [13] I. de la Torre-Diez, B. Martinez-Perez, M. Lopez-Coronado, J. R. Diaz, and M. M. Lopez, "Decision Support Systems and Applications in Ophthalmology: Literature and Commercial Review Focused on Mobile Apps," *J. Med. Syst.*, vol. 39, no. 1, 2014.
- [14] C. P. Wilkinson, F. L. Ferris, R. E. Klein, P. P. Lee, C. D. Agardh, M. Davis, D. Dills, A. Kampik, R. Pararajasegaram, J. T. Verdager, and F. Lum, "Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales," *Ophthalmology*, vol. 110, no. 9, pp. 1677–1682, 2003.