SHORT THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PhD)

Image processing and tear fluid proteomics based methods in the screening of diabetic retinopathy

by Dr. Zsolt Torok

Supervisor: Dr. Adrienne Csutak

UNIVERSITY OF DEBRECEN
DOCTORAL SCHOOL OF CLINICAL MEDICINE
DEBRECEN, 2015
Image processing and tear fluid proteomics based methods in the screening of diabetic retinopathy

By Zsolt Török MD

Supervisor: Adrienne Csutak MD, PhD
Doctoral School of Clinical Medicine, University of Debrecen

Head of the Examination Committee: Prof. Zoltán Hernádi MD, DSc
Members of the Examination Committee: Prof. János Németh MD, DSc
Prof. György Paragh MD, DSc

The Examination takes place at the Lecture Hall of the Department of Obstetrics and Gynecology, Faculty of Medicine, University of Debrecen, on March 12, 2015, at 11 AM.

Head of the Defense Committee: Prof. Zoltán Hernádi MD, DSc
Reviewers: Zsuzsanna Bereczky MD, PhD
Vanda Hidasi MD, PhD

Members of the Defense Committee: Prof. János Németh MD, DSc
Prof. György Paragh MD, DSc

The Ph.D. Defense takes place at the Lecture Hall of Building A, Department of Internal Medicine, Faculty of Medicine, University of Debrecen, on March 12, 2015, at 1 PM
1. Introduction

Diabetic retinopathy (DR) is one of the most common complications of diabetes mellitus (DM), accounting for about 5% of world blindness. Our work aimed to develop novel automatized methods for diabetic retinopathy (DR) screening.

Approximately 80% of all patients with at least 10 years of DM duration suffer from some degree of DR. During the development of DR, patients may not notice changes in their vision and DR might be very advanced by the time patients have visual complaints and experience visual loss eventually.

In order to detect DR in an early stage, everyone with DM should be subjected to a comprehensive dilated eye exam at least once a year. In case of early diagnosis, the progression of DR can be slowed down by appropriate systemic or local therapy.

Involvement of human graders is currently universal in the screening process and lower the automation in the system, the greater the costs. In the centralized screening programs, digital fundus images are captured at the place of patient care and forwarded to a grading center for evaluation by specially trained human graders or ophthalmologists. The system operates with high accuracy but due to its labor intensive nature, it might be poorly scalable in
economically challenged countries. In order to improve scalability and cost-effectiveness, many research groups are working on developing automated image analysis technologies.

During the study, our aims were the following:

- **Examination of the changes of tear fluid protein concentrations, caused by DR, using proteomics technics**
- **Identification of diabetic lesions on digitalized fundus images using image processing methods**
- **Combining the two of the above mentioned methods.**

**Diabetic retinopathy**

It is estimated that today more than 347 million people suffer from diabetes mellitus (DM) globally. Prevalence of diabetes in all age-groups was estimated to be 2.8% in 2000 and 4.4% in 2030. More than two-thirds of patients with DM live in developing countries where access to a high quality health care system is not universal. The meta-analysis of 35 studies, conducted between 1980 and 2008 provided data from 22,896 individuals with diabetes. The overall prevalence for any DR was 34.6% and 6.96% for proliferative DR. In the age group 40-60, the leading causes of blindness were myopia and DR, and macular degeneration together with DR in the older (above 60) population.
Diabetic retinopathy has two stages: i.) nonproliferative retinopathy; ii.) proliferative retinopathy. The progression of DR is accompanied by the development of visible retinal lesions. The hallmark of DR is microaneurysms (MAs). As these out-pouches occur on small blood vessels, most of the image processing-based screening methods concentrate on the detection of this type of lesion.

From image processing point of view, besides the MAs, the most characteristic retinal lesions are the hard exudates that are pale fatty deposits, which is a sign of leaking blood vessels. In the later stages, larger haemorrhages occur with new, abnormal, and fragile blood vessels that grow along the retina.

**Current screening methods**

The effectiveness of different screening modalities have been widely investigated. UK studies show sensitivity levels for the detection of sight-threatening diabetic retinopathy of 41-67% for general practitioners, 48-82% for optometrists, 65% for ophthalmologists, and 27-67% for diabetologists, and hospital physicians using direct ophthalmoscopy. Sensitivity for the detection of referable retinopathy by optometrists have been found to be 77-100%, with a specificity of 94-100%.

Regular DR screening is centralized in several developed countries
due to cost-efficiency and quality-control issues. Digital fundus images are captured at the place of patient care and forwarded to a grading center for evaluation by specially trained human graders or ophthalmologists. The process is performed with high accuracy. Sensitivity for the detection of sight-threatening diabetic retinopathy have been found 87-100% for a variety of trained personnel reading mydriatic 45° retinal photographs, with specificities of 83-96%. The British Diabetic Association (Diabetes UK) has established standard values for any diabetic retinopathy screening programme of at least 80% sensitivity and 95% specificity.

Based on the statement of the Health Technology Board for Scotland, the sensitivity, and specificity of slit lamp indirect ophthalmoscopy is an appropriate method for DR screening. Nevertheless, because of the lack of photo documentation, the quality assessment of the process is not feasible. The sensitivity and specificity of direct ophthalmoscopy is too low so the method is not eligible for DR screening purposes.

**Recent developments for DR screening**

In our current work we studied i.) Tear fluid proteomics and; ii.) Digitalized fundus imageprocessing-based methods.

Reports confirm *proteomics profile changes* in tear fluid under
irregular physiological and pathological conditions like wound healing or inflammatory diseases. Besides the proteins secreted by the lacrimal glands, tear fluid might contain proteins from epithelial cells covering the eye surface. Furthermore, proteins normally residing in the blood can get into the tear fluid through increased permeability of the conjunctival vessels. According to recent publications, the protein composition of tear fluid reflects normal or abnormal conditions, which justifies the use of tear fluid for screening purposes. Electrophoresis and chromatography also support that protein patterns of healthy and diabetic people are significantly different. Recently, high sensitivity and high resolution procedures were used for the assessment of the effects of trauma and various diseases in tear fluid of microliter quantity. In these cases, protein profiles were characterized by mass spectrometers MALDI-TOF, SELDI-TOF and LC/MS.

Several possible biomarkers for diabetic retinopathy were suggested. It was demonstrated using immunohistochemical methods, analysis of mRNA levels by qPCR or concentration measurement of individual proteins that serum levels of retinol binding protein 4, plasma levels of apelin, platelet-derived growth factor and vascular endothelial growth factor levels in vitreous and serum, lipopolysaccharide-binding protein and soluble CD14 levels in tear are elevated in case of patients with proliferative diabetic retinopathy, making
them potential biomarkers for the disease. These proteins were examined by targeted analysis without giving insights into global protein changes. As a broader approach, the protein composition of pooled protein extracts from neovascular membrane and non-vascular epiretinal membrane from patients with proliferative diabetic retinopathy were inspected by LC-MS/MS methods and it was found that periostin and pigment epithelium derived factor levels are significantly higher in the neovascular membrane. Kim, and his colleagues established a panel of blood plasma proteins serving as biomarkers for diabetic retinopathy using quantitative mass-spectrometry methods.

Our research group conducted as a first attempt to examine the use of tear fluid proteomics for DR screening by applying machine learning technics.

**The image processing-based methods** recognise the DR specific lesions on digital retina images using machine learning algorithms.

The most frequently used algorithms are created to detect the MA lesion. During the detection process, the software identifies the lesions that are presumably MA, and subsequently categorizes the images as healthy or diseased (DR and non-DR, respectively).

The digital image processing begins with a preprocessing step. Some fundus lesions are more visible after red filtering. Another important preprocessing step besides choosing the right color channels is standardizing
the size, intensity distribution and noise filtering of the images. It is followed by the application of algorithms that recognize healthy (e.g. vascular network, optic disc) and diseased features (usually small red or yellow/white lesions). The cutting edge methods for detecting lesions of the retina are based on learning algorithms. MAs are easy to define, relatively consequent in appearance and a sensitive early indicator for DR-onset, therefore the method presented by us is based solely on MA detection.

We combined tear proteomics-based and imageprocessing-based methods in our project, which is unlike other studies in this field.

2. Aims of the research

Examination of the global protein pattern changes in DR. Evaluation of the potential performance of the method, as well as the effect of the decrease in marker number on the sensitivity and specificity values. Improve performance of individual algorithms using the advantageous effect of integration of datasets coming from different data sources.

In the first phase, we focused on the applicability of tear fluid proteins as DR-markers. Our preliminary experiments successfully uncovered proteins that change their concentration in DR, and can therefore be used as markers. Our objectives were the following:
**Objective I:** After observing the unique concentrations of tear fluid proteins, we simultaneously measured the concentration of large number of proteins to identify global patterns that are typical to DR. According to our hypothesis, if we are able to examine the DR-induced variation of several proteins simultaneously, we can perform more precise classification of the screened population.

**Objective II:** After the examination of tear protein concentration patterns, we aimed to reduce the number of the proteins examined without decreasing the diagnostic sensitivity of the method. By excluding markers of low predictive values, increased cost-efficiency as well as simplified and faster feasibility can be achieved. This step is supposed to facilitate the application of the method in the clinical routine. Our long term goal is to develop a rapid test for detecting DR-presence from tear fluid, which can be applied in outpatient care.

**In the second phase,** we develop a DR-screening process by combining the tear fluid proteomics and image-processing-based methodologies. We formulated the following objective:

**Objective III:** To be able to compare our data to the current state of the art, we implemented an image processing method developed and published by others. The process was based on the detection of MAs that is the hallmark
of DR. After reproducing the result, we aimed to prove that the performance of the combined methodology exceeds i.) the performance of the method based solely on tear fluid proteomics ii.) as well as the performance of the image processing method.

3. Methods

Work phase I. – tear proteomics analysis

Patient examination

119 patients with DM were enrolled into the study from the Ophthalmology Outpatient Clinic of the University of Debrecen (HU). In the cases of 73 patients, one eye was examined due to difficulties in tear fluid sampling (e.g. keratoconjunctivitis sicca, conjunctivitis, operated eye, noncompliance) or because the retinal image could not be taken (e.g. cataract, hemorrhage, angle-closure glaucoma) or assessed (technical difficulties). Diabetic retinopathy was diagnosed by capturing 7-field fundus images of the patients, each evaluated by two independent ophthalmologists. Tear fluid samples were collected from the examined eyes by a trained assistant under standardized conditions. Tear samples were collected with glass capillaries immediately before pupil dilatation for fundus examination. Care was taken not to touch the conjunctiva.
Tear fluid proteomics examinations were done by the Department of Biochemistry and Molecular Biology, Proteomics Core Facility, University of Debrecen. Tear samples were examined using state-of-the-art nano-HPLC coupled ESI-MS/MS mass spectrometry protein identification. Global pattern of protein concentrations and its changes were described by the examination of concentrations of 34 different proteins.

By using machine learning algorithms we intended to predict the possible fulfillment of certain future events–using empirical data containing incomplete information. In the present case, input data (features) are protein levels measured in tear fluids from patients with diabetes and clinical data related to their DR status.

As the method is designed to be used for screening in the future, high sensitivity values were considered as the primary criteria for the construction of the classifier. This was aimed at minimizing the number of false-negative cases, thus reducing the chance for missing sight-threatening disease. The goal of this machine learning approach is to assess the best classification performance achievable by fitting different models to the dataset (model selection). In order to reach the best possible results, we tested sex classifiers algorithms: support vector machine (SVM), recursive partitioning (rpart), random Forest, naïve Bayes, logistic regression (logreg), k-nearest
neighbor(kNN). During the data analysis process, the R statistical framework has been used.

In order to monitor the classifier’s performance on the dataset, the testing was accomplished with \( k \)-fold cross-validation procedure. During the \( k \)-fold cross-validation the data set was divided into \( k \) equal parts. The first \( k-1 \) set (training set) is used for model construction and the later was tested on the \( k \)-fold set (test set). In the further \( k-1 \) iteration, the same procedure was followed, on the first \( k-2 \) and \( k \)-fold set, the model was learning and on the \( k-1 \), it was validated.

We ran cross-validation using the three different approaches below in order to find the best combination of input data and the different models.

First, we used data from all 34 identified proteins for model development.

However, we also wanted to analyze the changes in the performance of the classifiers if we only used a subset of the data. Second, only 6 marker proteins (out of the 34)–previously extracted by classical statistical methods–were used for decision making, applying the six machine learning algorithms. Third, we wanted to further evaluate the performance of the models by reducing the number of input variables, therefore we applied principal
component analysis (PCA). In this case, we performed dimension reduction to compress the information included in the original dataset with principal component analysis and used it for the visualization of the complex dataset in a 2D plane. With this method, the number of input features could be decreased while as much of the variation in data as possible could be retained.

**Work phase II. – combined method using data coming from tear fluid proteomics and image processing experiments**

**Patient examination:** 52 patients with DM were recruited into the project. Of all patients, 39 had signs of DR (24 non-proliferative, 15 proliferative). In our study, we only involved eyes of patients from who we could obtain complete tear fluid proteomics data and clinically-evaluable fundus photos. Out of the potential 104 eyes, 74 had corresponding tear fluid analysis and gradable digital retinal images. Digital retina images and tear fluid samples were taken from each eye at the same time using the method described above. In the proteomics data analysis, we used the concentration values of the 34 proteins. The main difference in this method compared to the method in Work Phase I, was that we added the results of the image processing examinations (numbers of MAs identified on the fundus images) to the protein concentration values as the input of the screening system.

The fundus images were processed using the MA detector algorithm,
and a threshold was chosen for a candidate to be classified as MA. The count of MA’s on an image found this way was used as the only feature for detecting DR, in the case of the first model.

**The MA-detector:** In our work, we implemented a MA detector published elsewhere in order to improve the comparability and reproducibility of our results. The green plane of the image was shade corrected, by subtracting the median filtered version of the image from the green component of each pixel value. On the shade corrected image, a contrast limited histogram equalization (CLAHE) was performed, which is used to enhance the contrast of the image.

This was followed by a 3x3 median filtering for smoothing. The next step of the processing is a top hat transform - an image processing method used for small feature extraction by morphological reconstruction. The reconstructed image is opened by a 10x10 disk shape to detect only small circular objects, resulting in candidate MAs. GBM model was used to classify candidates MAs based on the above features. The testing was done based on hand marked MAs of the retina images. In the next step, we used the same marked images to measure the performance of the MA detector.

Three classifiers were trained on the datasets; the first model was based on fundus image data alone, the second on using proteomics data, and
finally the combined model. During the process we used the best performing machine learning algorithms, Naive Bayes Classifier in the first and gradient boosting machine (GBM) in the other two models. 10 fold cross validation (repeated 10 times) was used to assess the performance of this classifier.

**Application of machine learning algorithm in the combined model**

**Learning phase:** We randomly select a sub-population of the total patient group, called the training group, and then using the known clinical diagnosis, split the training group into a DR and a non-DR group. The clinical diagnosis, the number of MAs on the retina images, and the protein concentration values are inputs of the machine learning algorithm. The algorithms are able to learn, which data patterns are the most characteristic for DR and non-DR groups.

**Assessment phase:** In the following steps, we use the data from the validation group. The number of MA-s and the protein concentration values constitute the input of the algorithm, but we do not use the information from clinical diagnosis. The learning algorithm compares the new data to the characteristic patterns that are known from the learning phase, and will make its own decision (normal/DR) for each patient as the output of the model. For the assessment of the performance of the model, we compare the output with the known clinical diagnosis.
4. Results

The results of work phase I. – tear fluid proteomics

Out of the six different machine learning algorithms, rpart methodology outperforms the other five models (0.74 sensitivity, 0.48 specificity and 0.65 accuracy). Naive Bayes method shows higher sensitivity (0.80) but low specificity (0.38). Random Forest method performs slightly worse than Recursive Partitioning (0.69 sensitivity, 0.45 specificity and 65% accuracy).

Six proteins were identified as independent biomarkers out of the 34 candidate proteins, by using statistical hypothesis testing. There is additional information in the joint distribution of the whole dataset, thus we wanted to examine the maximum performance of the model. Our reason for examining the performance of the reduced set was to develop a practical method for screening, without compromising the performance of the test. We have found that neither the models built on just the 6 marker proteins nor the models built on the PCA preprocessed data performed better or worse than the models built on the whole protein patterns. After PCA the first two components retained 22% of the variation of the original data set.
The results of work phase I. – tear fluid proteomics

In case of the first model MA-count was used as the only feature for detecting DR. *MA detection* method alone resulted in 0.84 sensitivity and 0.81 specificity values.

Using the *proteomics data* for analysis 0.87 sensitivity, and 0.68 specificity values were achieved by using GBM models.

The *combined model* resulted in a more powerful classifier, achieving 0.93 sensitivity and 0.78 specificity values, as the two different types of data provide independent and complementary aspects of the underlying information for the outcome.

5. Conclusions

Our aim was to develop a method that is able to pre-screen the patient population that suffer from DM. With the removal of negative cases (non-DR), the pre-screening process would cause the decrease of size of the patient population to be screened. Considering the high human resource demand of the current screening methods, including an automated step into the current protocol may increase the cost effectiveness of the screening program.
Over the last decade, several studies have been published on the application of image processing methods for DR screening. These methods, despite their promising performances, are only slowly being applied in clinical practice. Our team published the first attempt of using tear fluid proteomics complete with machine learning based method for DR screening. Although the conclusion of this work was that this method alone is not accurate enough for clinical use, there is a definite scope for the improvement of the performance of proteomics-based classifier. With improvements in tear fluid analysis, this method might become clinically useful with time, as it requires little equipment for obtaining the sample, making potential mass-screening of DR in hard-to-reach areas viable. In theory, the performance of a classifier might be improved by using the combination of different predictor models (image processing in our case).

The automated grading method published by Goatman, and his colleagues, intended to remove normal images from the image database before manual grading, thus reducing the manual workload. The combined method was based on the detection of MA, blot hemorrhages, and exudates. According to the well-known tradeoff between sensitivity and specificity, the 100% sensitivity is coupled with a low specificity value. Considering that the system proposed to be a pre-screening tool, the workload reduction ranging from
26.4% to 38.1% is remarkable.

Protein biomarkers and the MAs on retinal images represent different data sources and information on DR eye. Our results of 0.93 sensitivity and 0.78 specificity values are close to reach the threshold recommended for routine clinical screening of 80% and 95% respectively.

Our findings suggest that the maximal performance of this method has not been reached yet. Results can be improved in three potential ways: i) by optimizing the parameter settings of both tear fluid proteomics and MA detector based classifiers; ii.) by comparing and choosing the best classifier available for the combined screening method; iii.) by fine tuning the patient examination and lab diagnostic protocols.

Considering that we intend to develop a pre-screening method, a subsequent study on the assessment of potential workload reduction should also be done in order to decrease the workload of DR screening centers and maximizing our system sensitivity., Both tear sampling and retina photography are non-invasive methods and can be implemented at a general practitioner's settings. For the assessment of cost-effectiveness of the method, further analysis is needed. However, it is expected in the future that the cost of human resources in clinical care can become higher, in parallel with a rapid decrease in the cost of IT and laboratory technologies. In light of these
changes, the combined method of tear fluid proteomics and computer-assisted image processing of digital retinal photographs may provide a promising alternative in DR screening.

6. Summary

Diabetic retinopathy (DR) is one of the most common complications of diabetes mellitus (DM), accounting for about 5% of world blindness. Approximately 80% of all patients with type-2 DM duration of at least 10 years suffer from some degree of DR. During the development of DR, patients may not notice changes in their vision and DR might be very advanced by the time patients have visual complaints and experience visual loss eventually. Timely diagnosis and therapy, however, can significantly decelerate its progress, necessitating regular DR screening or appropriate follow-ups in all patients with DM. Involvement of human graders is currently universal and the lower the automation in the screening system, the greater the costs. Therefore, our aim was to develop a novel automated method for DR screening.

In the first phase of the project, our aim was to develop a novel method for DR screening based on the examination of tear fluid proteomics biomarker changes. We applied machine learning algorithms to predict
whether the given patient with DM suffers from DR or not, based on the
global proteomics pattern changes of their tear fluid. In the study we
concluded that, due to its low sensitivity (0.74) and specificity (0.48) values,
the proteomics-based screening method alone is not appropriate for clinical
application in its present format. However, in combination with other
methods, it is able to improve its overall performance.

In the second phase of the study, we used tear fluid proteomics based
methodologies with retinal image processing combined in one single system.
The imageprocessing-based method alone resulted in 0.84 sensitivity, and 0.81
specificity. Using the proteomics data for analysis 0.87 sensitivity, and 0.68
specificity values were achieved. The combined data analysis integrated the
features of the proteomics data along with the number of detected MAs in the
associated image, and achieved sensitivity/specificity values of 0.93/0.78. As
the two different types of data represent independent and complementary
information on the outcome, the combined model resulted in a reliable
screening method that is comparable to the requirements of DR screening
programs applied in clinical routine as specified by the British Diabetic
Association (at least 0.80 sensitivity, and 0.95 specificity).
6. New results

1. Work phase:

- (1.) In the project we successfully used machine learning algorithms in order to identified global patterns of individual tear fluid protein concentration that are specific to DR.

- (2.) We created a screening model that is, using the tear fluid protein concentration values, able to classify DM patients into DR and non-DR groups.

- (3.) Using the 6 marker proteins’ concentrations, previously identified by classical statistical methods, together we successfully created a screening method which is able to classify the DM patients into RD and non-RD groups.

2. Work phase:

- (4.) In an experimental model, we successfully combined tear fluid proteomics and image processing data for DR screening.

- (5.) The performance of the combined screening method exceeded the performance of the individual, proteomics or image processing methods.
7. Grant

This work was supported in part by the TAMOP-4.2.2.A-11/1/KONYV-2012-0045 Grant.
List of publications related to the dissertation

   DOI: http://dx.doi.org/10.1155/2014/623619
   IF:3.536 (2013)

   DOI: http://dx.doi.org/10.1186/1471-2415-13-40
   IF:1.075

   DOI: http://dx.doi.org/10.1016/j.jprot.2012.01.019
   IF:4.088
List of other publications


különböző fázisaiban.

**Total IF of journals (all publications): 12,233**
**Total IF of journals (publications related to the dissertation): 8,699**

The Candidate's publication data submitted to the IDEa Tudostér have been validated by DEENK on
the basis of Web of Science, Scopus and Journal Citation Report (Impact Factor) databases.

04 December, 2014