A classification model for predicting diabetic retinopathy based on patient characteristics and biochemical measures

Evangelia Kotsiliti[1,2], Bashir Al-Diri[1], Xujiong Ye[1], Maged Habib[3], Andrew Hunter[1]

[1]*School of Computer Science, University of Lincoln, Lincoln, UK*, [2]*Health Intelligence Ltd, Sandbach, UK*, [3]*Sunderland Eye Infirmary, Sunderland, UK*

Abstract

Purpose: In the United Kingdom (UK), the National Health Service (NHS) Diabetic Eye Screening Program offers an annual eye examination to all people with diabetes aged 12 and over aimed at the early detection of people at high risk of visual loss due to diabetic retinopathy. The purpose of this study was to design a model for predicting patients at risk of developing retinopathy using patient characteristics and clinical measurements.

Methods: We appraised data collected from the population-based Diabetic Eye Screening Program in East Anglia between 2011 and 2016. The data comprised retinal photographic screening results, patient characteristics of gender and age of the subject as well as duration and type of diabetes, and routine biochemical measures of Hemoglobin A1c (HbA1c), blood pressure, Albumin to Creatinine ratio (ACR), estimated Glomerular Filtration rate (eGFR), serum creatinine, cholesterol, and Body Mass Index (BMI). Individuals were classified according to the presence or absence of retinopathy as indicated by their retinal photographic screening results. A logistic regression with Least Absolute Shrinkage and Selection Operator (lasso) regularization, random forest, gradient boosting machine, and regularized gradient boosting model were deployed and cross-validated for their predictive ability.

Results: A total of 6375 subjects with recorded information for all available biochemical measures were identified from the cohorts. Of these, 5969 individuals had no signs of diabetic retinopathy. Of the remaining 406 individuals with signs of diabetic

---

Correspondence: Evangelia Kotsiliti, School of Computer Science, University of Lincoln, Lincoln, UK.
E-mail: ekotsiliti@lincoln.ac.uk

retinopathy, 352 had background diabetic retinopathy and 54 had referable diabetic retinopathy. The highest value of the ten-fold cross-validated area under the curve (AUC) of the receiver operating curve (ROC), 0.73 ± 0.03, was achieved by the gradient boosting machine and the minimum required set of variables to yield this performance included: duration of diabetes, HbA1c, ACR, and age. A subsequent analysis on the predictive power of the biochemical measures showed that when HbA1c and ACR measurements were available for longer time periods, the performance of the models was enhanced. When HbA1c and ACR measurements for a five-year period prior to the event of study were available, gradient boosting machine cross-validated AUC was 0.77 ± 0.04 in comparison to the cross-validated AUC of 0.68 ± 0.04 when only information for the one-year period for these variables was available. Similarly, an increment from 0.70 ± 0.02 to 0.75 ± 0.04 was observed with random forest. The dataset with the one-year measurements comprised 4857 subjects, of whom 4572 had no retinopathy and the remaining 285 had signs of retinopathy. The dataset with the five-year measurements comprised 757 subjects, of whom 696 had no retinopathy and the remaining 51 had signs of retinopathy.
Conclusion: Patient information and routine biochemical measures can be used to identify patients at risk of developing retinopathy with a significant reduction of the number of screening visits.

Keywords: area under the ROC curve (AUC), classification, gradient boosting, lasso, prevalence and risk of diabetic retinopathy, random forest, retinal screening

1. Introduction

Diabetic retinopathy and diabetic maculopathy are the most common microvascular complications of diabetes and among the leading causes of blindness in the UK.[1] Diabetic retinopathy accounts for about 7% of people who are registered blind in England and Wales.[2,3]

The Diabetic Eye Screening Program in the UK has been in place for over a decade and recommends an annual eye examination to all people diagnosed with diabetes aged 12 and over. The principal goal of the program is to prevent diabetes-related blindness, with individuals identified at high risk being referred to eye clinics for further assessment and management.[4] Since 1996, the number of people with diabetes in the UK has more than doubled from 1.4 million to 3.3 million.[2,5] In England alone, there are $\sim$ 2.8 million people living with diabetes,[2] the majority of whom (99%) are eligible for screening.[6]

Examination of the retinal grading outcomes of people attending the population-based Diabetic Eye Screening Service in East Anglia showed that the majority of people had no retinopathy. There are $\sim$ 90,000 people screened by the Diabetic Eye Screening Service in East Anglia on an annual basis, whose retinal photographic images are assessed through a quality-assured multi-level grading scheme compliant with national recommendations.[7] The prevalence of diabetic retinopathy was cal-

culated for the years 2011 to 2016 for all people with given consent and assessable retinal images. Table 1 shows the prevalence of diabetic retinopathy per year.

Table 1. Prevalence of diabetic retinopathy in the population-based Diabetic Eye Screening Program in East Anglia.

	2011* N = 28,042	2012 N = 68,447	2013 N = 75,417	2014 N = 82,209	2015 N = 83,767	2016* N = 21,931
No retinopathy	0.70	0.70	0.72	0.71	0.71	0.70
Background retinopathy	0.26	0.26	0.24	0.26	0.26	0.26
Referable retinopathy	0.04	0.04	0.04	0.03	0.03	0.04

*For the years 2011 and 2016, information was available for the second half of 2011 and the first trimester of 2016.

Over the years, 70-72% of the population screened annually were graded with no retinopathy, 24-26% were graded with background retinopathy, and 3-4% were graded with referable retinopathy. Similar prevalence of diabetic retinopathy is also reported from other screening services across the UK.[8–10] These figures imply that an annual recall for screening of all patients with diabetes may not be the optimal approach.

In this study, we used longitudinal data of patient characteristics and biochemical measures from the population-based Diabetic Eye Screening Service in East Anglia to build a model for predicting the presence/absence of diabetic retinopathy. Effective identification has the potential to save on patients' traveling costs and time without compromising their health. It can additionally reduce the workload of screening services and save on healthcare resources, something which has become of utmost importance due to the ever-increasing number of patients with diabetes.[11]

2. Methods

Data comprised retinal photographic screening results, patient characteristics of gender and age of the subject as well as duration and type of diabetes, and routine biochemical measures of Hemoglobin A1c (HbA1c), blood pressure, Albumin to Creatinine ratio (ACR), estimated Glomerular Filtration rate (eGFR), serum creatinine, cholesterol, and Body Mass Index (BMI).

Retinal photographic screening results and patient characteristics were extracted in an anonymized manner from the population-based Diabetic Eye Screening Program in East Anglia between 01/07/2011 and 31/03/2016 for subjects with given consent. Subjects invited to the retinal screening program include all people with diabetes aged $\geq$ 12 years. There were 104,366 subjects who attended the service within that period. The median number of screening episodes was 4 (interquartile range 2-5). At each screening episode, two color retinal images of standard 45° fields (macula and disc centered) were taken per eye after pupil dilatation. The retinal images were subsequently assessed by accredited personnel who assigned grades for the presence and severity of diabetic retinopathy via a quality-assured multi-level scheme that meets national standards.[7]

Grades assigned include R0 for no retinopathy, R1 for background retinopathy, R2 for non-proliferative retinopathy, R3a for active proliferative retinopathy, R3s for stable proliferative retinopathy, and M1 for maculopathy. Subjects with R0, R3s, and R1 are invited for annual re-screen, and in the latter case, the General Practice is also notified so that these people can be advised of appropriate actions to limit risk of disease progression. Subjects with R2, R3a, and/or M1 (hereafter referred to as referable grades) are referred to ophthalmology clinics for further assessment and management. Detailed information regarding the feature-based grading criteria and relationship to the Early Treatment Diabetic Retinopathy Study (ETDRS) severity scale can be found on the Diabetic Eye Screening Program website.[12]

For the subjects who attended the screening service within that period, measurements of different biochemical measures were also extracted. These measurements were taken for each subject during their routine checks at the GP practices as part of their diabetes care. The number of measurements and the time period over which they were taken varied among subjects. The annual number of measurements ranged from one to four, and the time horizon over which they were taken extended from one up to ten years prior to the last retinal screening appointment. Therefore, the multiple measurements of each biochemical measure were averaged before they were inserted into the models.

Subjects were categorized into two groups, no retinopathy (NR) and any retinopathy (AR), based on the results from the retinal photographic images (retinal grades) and after meeting certain requirements.

Subjects included in the NR group had at least three eye examinations resulting in R0. The minimum number of eye examinations was set at three to minimize the impact of possible misclassification errors in the grading process. For example, a subject with a grading pattern of 'R0 - R0 - R0', increased our confidence that the subject was indeed free of retinopathy during the second examination in comparison to a subject with a pattern of 'R0 - R0 - R1'. The last grade was only used as a reassurance grade for the absence of retinopathy, and therefore, the clinical profile of included subjects was up to the screening date before the last. This requirement has led to the exclusion of subjects with transient patterns between R0 and R1 (*i.e.*, R0 – R1 – R0). Longitudinal examination of the grading outcomes in the data showed a consid-

erable number of subjects with such transient patterns. For those of short grading history, *e.g.*, two or three grading outcomes, it was impossible to conclude whether these patterns resulted from over-grading or under-grading. Possible factors to such patterns are the variability of human performance, confusion of photographic artifacts, and media opacity and sensitivity of fundus photography to depict changes occurring deep in the retina. Similar observations were also reported from the Scottish Diabetic Eye Screening Program,[13] where 18.2% of people with type 1 diabetes and 36.4% of people with type 2 diabetes transitioned from mild background retinopathy to no visible retinopathy at their subsequent examination.

Subjects included in the AR group met any of the following two sets of requirements.

1. *Set 1:* At least four eye examinations of the pattern 'R0 - R0 - ... - R1 - R1 or worse'. This set aims to capture the subjects with onset of retinopathy (R1). The subsequent grading outcome after R1 serves only as a reassurance grade, and therefore, the clinical profile of the included subjects was up to the first R1 event. The reassurance grade needs to be R1 or worse, and therefore subjects with a pattern of 'R0 - R0 - ... - R1 - R0' were excluded, as in that case, it would be impossible to judge whether subjects were under-graded at the last event or over-graded at the event prior to the last.

2. *Set 2:* At least three eye examinations, the last of which resulted to a referable grade and all grades prior to that were R0. Here, no reassurance grade was required because: a) referable grades were assigned by the graders at the top level of the hierarchy, who we assumed to have optimal performance; and b) a subsequent grade of R0 could be attributable to subjects possibly being treated after referral to the hospital. Subjects with evidence of previous laser treatment or unassessable images were excluded from both groups.

To utilize all available biochemical measures, only subjects with a complete clinical profile were included in the study. This resulted in 5969 subjects in the NR group and 406 subjects in the AR group. Of the 406 subjects in the AR group, 352 had R1 and the other 54 had referable diabetic retinopathy. The characteristics of the population included in the study are shown in Table 2 and Table 3.

Table 2. Patient characteristics (categorical variables) of the 5969 subjects in the NR group and 406 subjects included in the AR group.

	NR = 5969	AR = 406
Gender (male)	57%	51%*
Diabetes Type (type 2)	96%	95%

*Statistical significance ($p < 0.05$). Tested with the Chi-squared test.

The transformed features that were extracted from the various biochemical measurements were considered for the time period prior to the event of study, and the

Table 3. Patient characteristics (continuous variables) of the 5,969 subjects in the NR group and 406 subjects included in the AR group.

	Median (25th 75th centile)	
	NR = 5969	**AR = 406**
Age (years)	68 (60-76)	70 (59 78)*
Duration of diabetes (years)	7 (3-11)	9 (6-13)*
HbA1c (mmol/mol)	54 (48-61)	59 (53-68)*
Systolic blood pressure (mmHg)	133 (128-138)	133 (128-138)*
Cholesterol (mmol/L)	4.6 (4.0-5.1)	4.5 (4.0-5.0)*
Serum Creatinine (mmol/L)	86 (76-98)	85 (75-97)*
ACR (mg/mmol)	1.01 (0.9-1.8)	1.40 (1.0-2.8)*
eGFR (mL/min/1.73m^2)	78 (67-88)	79 (68-87)*
BMI (kg/m^2)	30.5 (27.1-34.6)	30.2 (26.9-34.5)*

*Statistical significance ($p < 0.05$). Tested with the Mann-Whitney U test.

features related to the patient characteristics were considered at the date of the event of study. These features were investigated for their ability to separate between NR and AR.

Both linear and non-linear techniques were deployed and compared for their predictive ability. These comprised logistic regression with lasso regularization,[14,15] random forest,[16] gradient boosting machine,[17] and extreme or regularized gradient boosting machine.[18]

Lasso regularized logistic regression has shown good performance in the presence of many irrelevant features,[19] and can perform feature selection via the lasso penalty term. Selecting a good value for the lasso penalty is critical and the process for selecting it is described later in this section.

Random forest, gradient boosting machine, and extreme or regularized gradient boosting machine are decision tree ensemble models which involve a combination of a large number of trees. In a random forest, several decision trees are built on bootstrapped training samples, and each time a split in a tree is considered, only a random subset of the predictors is chosen as split candidates from the full set of predictors. This process leads to the decorrelation of the built trees, thereby making the average of the resulting trees less variable and hence more reliable. In a gradient boosting machine, the trees are grown sequentially. In more detail, each tree is grown using information from previously grown trees in order to minimize a loss function. A regularization term is also involved and controls the rate at which boosting learns. In general, statistical learning approaches that learn slowly tend to perform well. Extreme or

regularized gradient boosting machine works in a similar way as the gradient boosting machine, but there are differences in the modeling details of the regularization term. These tree ensembles can often result in dramatic improvements in prediction accuracy over a single decision tree, and were deployed due to the differences in the underlying modeling principles and the varying number of parameters involved, the tuning of which could have a considerable impact on the predictive performance.

The predictive performance of each model was assessed by the AUC value on unseen data. To create the unseen data from the existing set of observations, a ten-fold cross validation approach was deployed that suffers neither from excessively high bias nor from very high variance in comparison to the validation and leave-one-out approaches.[20]

Parameter tuning was performed during the model fitting phase with the use of five-fold cross validation. In more detail, the train set was further divided into five folds, one of which was treated as a test set, and the remainder as train set. To avoid confusion, we refer to these sub-divided sets as internal train and test sets. The choice of five folds was due to the limited number of training samples. For each internal train set, a grid of values for the tuning parameter(s) was supplied, and for each of those values or combination of them, the model was fitted. Subsequently, the model was assessed on the internal test set by estimating the AUC. The value or combination of values of the tuning parameter(s) that maximized the internal test set AUC rendered a candidate value. The same process was repeated across all folds, and the candidate values that were chosen most frequently formed the final optimal values.

In the lasso regularized logistic regression, variable selection is inherently performed during parameter tuning, due to the fact that the parameter has the effect of shrinking the coefficient estimated values towards zero and forcing some of them to be exactly equal to zero when a sufficiently large value is selected. In the case of the tree-based models, a backward stepwise variable selection approach was followed. Initially, the models were trained on the full set of features and their predictive ability was assessed on the test set by estimating the AUC. At the next step, the variable with the lowest mean decrease in Gini index (a measure of node purity) was eliminated and the model was trained and assessed on the test set for the remainder set of predictors. The process was repeated until only one feature remained. All analyses were performed in R Statistical Computing software.[21]

Given the fact that the number of measurements for the biochemical measures as well as the time window over which they were taken varied among subjects, we subsequently conducted an analysis to investigate whether the utilization of clinical information for shorter and longer periods of time would impact the performance of the models. Initially, univariate logistic regression was deployed to assess the predictive power of each biochemical measure at different time windows. The averaged value of the measurements at each observational time window was used as input in the model. Time windows ranged from one up to ten years before the event of study. The predictive power was assessed in the context of ten-fold cross-validated AUC.

The best independent predictors, *i.e.*, those that yielded the highest cross-validated

AUC (and above 0.5), were HbA1c and ACR. At the subsequent step, these predictors were used to form two datasets. Both datasets included the entire set of predictors, but in one dataset the measurements of HbA1c and ACR were only available for the one-year period prior to the event of study, whereas in the other the measurements of HbA1c and ACR were available for the five-year period prior to the event of study. The five-year period was the maximum period of time we could consider to form the second dataset, since, after this year, only a few subjects had a full clinical record of the entire set of predictors. The dataset with the one-year measurements comprised 4857 subjects, of whom 4572 had no retinopathy, while the remaining 285 had signs of retinopathy. The dataset with the five-year measurements comprised 757 subjects, of whom 696 had no retinopathy and the remaining 51 had signs of retinopathy.

Gradient boosting machine and random forest were deployed in both datasets to quantify the impact that the exploitation of HbA1c and ACR information for longer periods of time could have on the models' predictive ability. Gradient boosting machine was deployed due to the fact that it yielded the best predictive performance in the previous experiment, and the latter in order to investigate whether there was a consistency in the outcomes between the two learning techniques.

Given the small number of subjects belonging to the retinopathy group, particularly in the five-year dataset, five-fold cross validation was used instead of ten-fold cross validation in this experiment to assess predictive performance. Parameter tuning and feature selection were performed in the same way as described previously.

3. Results

Logistic regression with lasso regularization, random forest, gradient boosting machine, and extreme or regularized gradient boosting machine were fitted on the dataset comprising 5969 individuals with no signs of diabetic retinopathy and 406 individuals with signs of diabetic retinopathy. For these individuals, measurements for the entire set of biochemical measures were available. The available measurements were averaged prior to inclusion in the models; therefore, the entire set of features/predictors comprised 11 variables: the mean value of measurements for each biochemical measure and the value for each of the patient's characteristics.

Gradient boosting machine had the best performance with 0.73 for the cross-validated AUC and standard deviation of 0.03 (0.73 ± 0.03). The minimum subset of variables to yield that performance included four variables: diabetes duration, mean value of HbA1c measurements, mean value of ACR measurements, and subject age. The performance of the model was stabilized to AUC 0.73 ± 0.03 for all subsequent subsets of variables. The model of four variables yielded an AUC increment of 0.01 and lower standard deviation over the model of three variables, which comprised duration of diabetes, HbA1c, and ACR.

Random forest and regularized gradient boosting yielded an AUC of 0.72 ± 0.05 and 0.72 ± 0.04, respectively. The best performance was achieved for the same subset

of variables as in gradient boosting, and it remained the same or slightly changed (± 0.01) for the subsequent sets of variables.

The lasso regularized logistic regression had an inferior performance over the rest of the models with an AUC 0.70 ± 0.02. In all ten folds, the lasso regularized logistic regression consistently selected a set of three variables: duration of diabetes, mean value of HbA1c measurements, and mean value of ACR measurements. The coefficients of all predictor variables were positive, meaning that subjects with higher mean values of HbA1c and ACR as well as longer duration of diabetes are more likely to develop diabetic retinopathy. These findings of HbA1c and duration of diabetes are in concordance with the findings of major epidemiological studies.[22–24]

Figure 1 displays the mean value of the ten-fold cross-validated AUC for the random forest, gradient boosting machine, and regularized gradient boosting models by the number of variables which remained in the model at each step of backward feature selection.
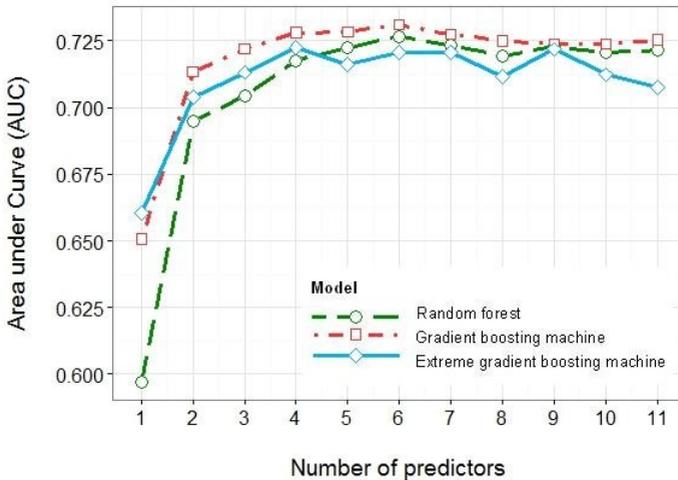
Fig. 1. Mean value of the ten-fold cross-validated AUC for random forest, gradient boosting machine, and regularized gradient boosting machine for different sets of variables formed by stepwise backward selection.

Subsequently, univariate logistic regression was deployed to compare the independent predictive power of each biochemical measure for time-varying observational windows. The observational time windows ranged from one up to a maximum of ten years prior to the event of study (measurements of some biochemical measures were available for longer time periods than others). The predictive power was assessed in the context of ten-fold cross-validated AUC, and the results are represented in Table 4.

As can be seen in Table 4, the best independent predictors for all biochemical measures, *i.e.*, those that yielded the highest value of AUC, were HbA1c and ACR. In

more detail, the ten-fold cross-validated AUC for HbA1c gradually increased from 0.62 to 0.66 for longer time periods, and a gradual increment from 0.60 to 0.68 was also observed in the case of ACR. The predictive power of the remaining variables was poor, *i.e.*, AUC $\sim$ 0.50, and most of them did not seem to be affected by the varying time windows. An inverse effect of longer time windows on the predictive power of systolic blood pressure and cholesterol was observed. In particular, the ten-fold cross-validated AUC for systolic blood pressure gradually decreased from 0.54 to 0.51 for longer time periods; a gradual decrement from 0.55 to 0.43 was also observed in the case of cholesterol.

Since HbA1c and ACR were the most significant independent clinical predictors, showing improved predictive power for longer observational time windows, models were subsequently re-fitted into two newly formed datasets in which the measurements of these two predictors were available for a one-year and five-year period of time, respectively. The remaining clinical predictors were also included in these datasets, but the extent of their measurements was not constrained to a particular limit of time. The dataset with the one-year measurements comprised 4857 subjects, of whom 4572 had no retinopathy and the remaining 285 had signs of retinopathy. The dataset with the five-year measurements comprised 757 subjects, of whom 696 subjects had no retinopathy and the remaining 51 had signs of retinopathy.

Gradient boosting machine and random forest were fitted in both datasets and the results of the five-fold cross-validated AUC are displayed in Figure 2. In both models, predictive ability was improved for the five-year dataset in comparison to the one-year dataset. In particular, gradient boosting machine yielded an increment of the five-fold cross-validated AUC from 0.68 ± 0.04 for the one-year dataset to 0.77 ± 0.04 for the five-year dataset. Similarly, random forest yielded an increment of the five-fold cross-validated AUC from 0.70 ± 0.02 for the one-year dataset to 0.75 ± 0.04 for the five-year dataset.
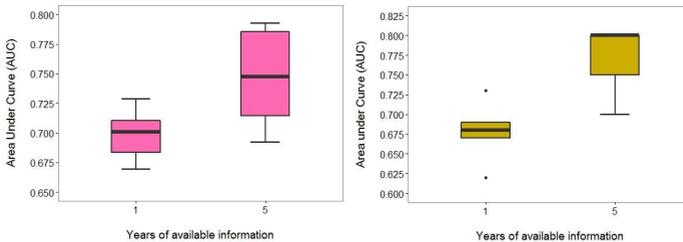
Fig. 2. The five-fold values of AUC for the one-year and year-year datasets. The figure on the left shows the values of AUC yielded by random forest; the one on the right shows the values of AUC yielded by gradient boosting machine.

Table 4. Results of the univariate logistic regression analysis at time-varying observational windows. The predictive power of independent predictors is expressed by the value of the ten-fold cross-validated AUC (mean ± standard deviation).

Time window (in years)	Biochemical measures						
	HbA1c	ACR	Systolic BP[*]	Cholesterol	eGFR	Creatinine	BMI
1	0.62 ± 0.03	0.60 ± 0.05	0.54 ± 0.03	0.55 ± 0.04	0.51 ± 0.03	0.51 ± 0.02	0.51 ± 0.03
2	0.62 ± 0.03	0.60 ± 0.07	0.54 ± 0.03	0.55 ± 0.05	0.52 ± 0.03	0.52 ± 0.01	0.52 ± 0.03
3	0.63 ± 0.03	0.63 ± 0.05	0.54 ± 0.05	0.53 ± 0.06	0.51 ± 0.04	0.51 ± 0.02	0.52 ± 0.03
4	0.64 ± 0.02	0.65 ± 0.12	0.53 ± 0.05	0.53 ± 0.05	0.50 ± 0.04	0.51 ± 0.02	0.52 ± 0.03
5	0.64 ± 0.03	0.66 ± 0.07	0.53 ± 0.05	0.53 ± 0.07	0.50 ± 0.03	0.51 ± 0.04	0.51 ± 0.05
6	0.64 ± 0.05	0.68 ± 0.15	0.53 ± 0.07	0.43 ± 0.07	0.47 ± 0.02	0.51 ± 0.04	0.51 ± 0.03
7	0.64 ± 0.05	NA[*]	0.52 ± 0.05	0.47 ± 0.10	0.49 ± 0.04	0.51 ± 0.03	0.51 ± 0.05
8	0.64 ± 0.05	NA[*]	0.52 ± 0.05	0.42 ± 0.04	0.51 ± 0.06	0.50 ± 0.03	0.51 ± 0.04
9	0.66 ± 0.04	NA[*]	0.52 ± 0.03	0.45 ± 0.08	0.54 ± 0.14	0.50 ± 0.05	0.51 ± 0.05
10	NA[*]	NA[*]	0.51 ± 0.04	0.43 ± 0.10	NA[*]	0.50 ± 0.05	0.51 ± 0.05

[*]NA indicates that the number of individuals with measurements up to that period of time was not adequate to perform ten-fold cross validation.

4. Discussion

This study explored the potential for predicting the risk of developing diabetic retinopathy with the use of patient characteristics and clinical biochemical measures. Of all models considered in the analysis, gradient boosting machine had the best predictive ability, yielding an AUC of 0.73 ± 0.03. In addition, all non-linear models yielded better predictive ability than the linear model, indicating that there could be a more complex underlying relationship between the predictors and the response that the linear model failed to capture.

The model that yielded the best predictive performance included the following variables: patient age, diabetes duration, mean value of HbA1c measurements, and mean value of ACR measurements. Epidemiological studies have reported the effect of age, duration of diabetes, and HbA1c on the incidence and progression of diabetic retinopathy.[24–28] The Diabetes Control and Complications Trial (DCCT)[24] and the United Kingdom Prospective Diabetes Study (UKPDS)[25] have demonstrated that improved blood glucose control reduced the risk of developing diabetic retinopathy. This is in concordance to our finding of improved predictive power of HbA1c for longer observational time windows, due to the fact that the extended measurements of HbA1c provide an indication of prolonged poor or good glycemic control. Our finding that ACR is a significant predictor for diabetic retinopathy is in agreement with the findings of Yanming *et al.*[28] that ACR plays an important role in the pathogenesis of diabetic retinopathy. Moreover, we found that the predictive ability of ACR was improved when measurements were available for longer periods of time prior to the event of study.

The remaining biochemical measures (systolic blood pressure, cholesterol, serum creatinine, eGFR, and BMI) were not included in our final model due to their poor predictive ability. Although associations of diabetic retinopathy with these variables have been documented by several studies,[25,29–36] findings were inconsistent and a definitive correlation between them and the incidence and progression of diabetic retinopathy has not been established. Even when longer measurement periods for these variables were considered in the present study, their predictive ability did not show any improvement. On the contrary, in the case of systolic blood pressure and cholesterol, an adverse effect was observed. A comprehensive review about blood pressure control for diabetic retinopathy by Do *et al.*[37] concluded that the available evidence supports a beneficial effect of intervention to reduce blood pressure with respect to preventing diabetic retinopathy for up to four to five years. Additionally, it stated that the existing evidence weakens the conclusion regarding an overall benefit of intervening on blood pressure solely to prevent diabetic retinopathy. These conclusions shed some light on the declining predictive ability of systolic blood pressure on the development of diabetic retinopathy for longer observational time windows that was observed in the present study. Systolic blood pressure yielded an AUC of 0.54 for the one-year up to the three-year observational time windows, and then gradually deteriorated reaching an AUC of 0.51 for the ten-year observational time window. The

deteriorating predictive ability of cholesterol for increasing time windows requires further exploration due to the number of factors that might affect cholesterol levels.

Other studies focused on the risk assessment of diabetic retinopathy were found in the literature.[38–42] Cichosz *et al.*[38] developed a model for the identification of patients at high risk of having retinopathy at the time they were diagnosed with type 2 diabetes. The model yielded an AUC of 0.74, but it required additional information on the following variables: BMI, waist circumference, systolic blood pressure, urinary albumin, and urinary creatinine. In addition, the model was validated on a small sample and the stability of predictive performance (*e.g.*, standard deviation) was not reported. Moreover, this model could only be applicable to patients with newly diagnosed type 2 diabetes.

The remaining studies had a different design aiming at the identification of patients at high risk of developing referable diabetic retinopathy and used additional information on the presence and severity of diabetic retinopathy at previous screening appointments. Therefore, these models could not be applicable to detect the onset of diabetic retinopathy and they additionally require information on past screening grading outcomes. Mehlsen *et al.*[39] presented a model with which, on average, the screening interval could be prolonged by approximately 2.9 times in patients with type 1 diabetes, and by 1.2 times in patients with type 2 diabetes. This model, apart from clinical information, also requires information on the recorded severity of diabetic retinopathy, which is expressed by the number of retinal hemorrhages and hard exudates counted in macula-centred 60° fundus photographs. Aspelund *et al.*[40,41] developed a model which required information on the presence and severity of retinopathy at the last screening appointment. This model yielded an AUC of 0.80 (CI 0.78 to 0.81) for patients with type 2 diabetes, and 0.70 (CI 0.67 to 0.73) for patients with type 1 diabetes. Scanlon *et al.*[42] presented a model which yielded an AUC of 0.80 and also required information on the presence and severity of diabetic retinopathy at the last screening appointment.

As these models were not directly comparable to the model in the present study due to their different scope, design, and presentation of results (*i.e.*, AUC was not reported in some cases), we decided to investigate how the model in our study performed on patients with R1 and referable retinopathy separately. We chose gradient boosting machine to assess performance and we selected an overall threshold for the model in order to categorize the subjects into one of the two groups. The classification results were aggregated for the ten folds. For a very small value of the threshold, 11 out of 352 subjects with R1 were misclassified as R0 (3% error) and 2 out of 54 with referable retinopathy (4% error) were misclassified as R0. This, however, decreased the specificity of the model to 20%, which, if considered for the 90,000 population in the East Anglia screening program, could still lead to ∼ 17,000 less screening visits. When a bigger threshold value was applied, specificity increased to 46% (∼ 40,000 less screening visits), 17 out of 352 subjects with R1 were misclassified as R0 (5% error rate) and 10 out of 54 with referable retinopathy (19% error rate) were misclassified as R0.

Threshold values, however, can be determined on the basis of acceptable error for the referable grades, which, based on the findings of major clinical trials, should be low. In brief, the findings from the large clinical trials are the following: the annual risk of progressing from R1 to R2 and/or R3 is $\sim$ 8% for people who receive no medical intervention,[24] the annual risk of progressing from R2 to R3 is 20% for people with type 2 diabetes if left untreated,[43] and untreated eyes with R3 characteristics have a risk of 15.9% $-$ 26% of progressing to severe visual loss within two years.[44] These figures suggest that patients with referable retinopathy have at least a 2.5-fold higher risk to vision loss/deterioration than patients with R1, and so misclassification error for these people shall be really low. An exception for the case of R1M1 could be applied by allowing a bigger margin of error in the face of the findings of the study[45] which reported that only 10% of the recorded M1 represent clinically significant macular edema requiring treatment.

We have shown how a model exploiting patient characteristics and biochemical measures can be used to identify patients at risk of developing retinopathy. Effective identification could reduce the number of screening visits and save on health care resources without jeopardizing patient care. As the numbers of people diagnosed with diabetes are rising, this can lead to challenges providing retinal screening. The model could be used by the General Practices, which patients with diabetes visit throughout the year for routine blood tests and, in the event of a positive risk assessment, direct them to the screening service.

5. Conclusions and future perspectives

In this work we presented a model for the identification of people with diabetes at risk of developing diabetic retinopathy. In the future, we aim to investigate whether additional information, *e.g.*, co-morbidities, ethnicity, and medication could enhance the predictive ability of the model.

Acknowledgements

References

1. Liew G, Michaelides M, Bunce C. A comparison of the causes of blindness certifications in England and Wales in working age adults (16–64 years), 1999–2000 with 2009–2010. BMJ open, 2014;4(2): e004015.
2. Diabetes U. Facts and stats. 2015; Available from: https://www.mrc.ac.uk/documents/pdf/diabetes-uk-facts-and-stats-june-2015.

3. Leamon S, Davies M. Number of adults and children certified with sight impairment and severe sight impairment in England and Wales: April 2011–March 2012. RNIGB and Moorfields Hospital NHS Foundation Trust, 2013;

4. Peto T, Tadros C. Screening for diabetic retinopathy and diabetic macular edema in the United Kingdom. Current diabetes reports, 2012;12(4): 338–345.

5. Curnow S. Review: Diabetic Association, BDA, Diabetes UK. The British Journal of Diabetes & Vascular Disease, 2002;2(6): 469–472.

6. England N. Integrated Performance Measures Monitoring:Diabetes Data. 2013; Available from: https://www.england.nhs.uk/statistics/statistical-work-areas/integrated-performance-measures-monitoring/diabetes-data/.

7. Team NC. Diabetic Eye Screening Pathway Overviews. 2012; Available from: https://www.gov.uk/government/publications/diabetic-eye-screening-pathways-patient-grading-referral-surveillance.

8. Scotland GS, McNamee P, Philip S, Fleming AD, Goatman KA, Prescott GJ, et al. Cost-effectiveness of implementing automated grading within the national screening programme for diabetic retinopathy in Scotland. British Journal of Ophthalmology, 2007;91(11): 1518–1523.

9. Looker H, Nyangoma S, Cromie D, Olson J, Leese G, Black M, et al. Rates of referable eye disease in the Scottish national diabetic retinopathy screening programme. British Journal of Ophthalmology, 2014;98(6): 790–795.

10. Thomas RL, Dunstan FD, Luzio SD, Chowdhury SR, North RV, Hale SL, et al. Prevalence of diabetic retinopathy within a national diabetic retinopathy screening service. British Journal of Ophthalmology, 2015;99(1): 64–68.

11. Federation ID. IDF Diabetes Atlas Sixth edition. 2014;

12. England PH. Diabetic eye screening: retinal image grading criteria. Available from: https://www.gov.uk/government/publications/diabetic-eye-screening-retinal-image-grading-criteria.

13. Looker H, Nyangoma S, Cromie D, Olson J, Leese G, Philip S, et al. Predicted impact of extending the screening interval for diabetic retinopathy: the Scottish Diabetic Retinopathy Screening programme. Diabetologia, 2013;56(8): 1716–1725.

14. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 1996; 267–288.

15. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. Journal of statistical software, 2010;33(1): 1.

16. Liaw A, Wiener M. Classification and regression by randomForest. R news, 2002;2(3): 18–22.

17. Friedman JH. Greedy function approximation: a gradient boosting machine. Annals of statistics, 2001; 1189–1232.

18. Chen T, He T. xgboost: eXtreme Gradient Boosting. R package version 0.4-2, 2015;

19. Lee S-I, Lee H, Abbeel P, Ng AY. Efficient l˜ 1 regularized logistic regression. Proceedings of the National Conference on Artificial Intelligence. 21. (1): Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999. 2006; 401.

20. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. 6. Springer, 2013; 184.

21. Team RC et al. R: A language and environment for statistical computing. 2015;

22. Klein R, Klein BE, Moss SE, Davis MD, DeMets DL. The Wisconsin Epidemiologic Study of Diabetic Retinopathy: II. Prevalence and risk of diabetic retinopathy when age at diagnosis is less than 30 years. Archives of ophthalmology, 1984;102(4): 520–526.

23. Group UPDSU et al. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). The Lancet, 1998;352(9131): 837–853.

24. Control D, Group CTR, et al. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl j Med, 1993;1993(329): 977–986.

25. Stratton I, Kohner E, Aldington S, Turner R, Holman R, Manley S, et al. UKPDS 50: risk factors for incidence and progression of retinopathy in Type II diabetes over 6 years from diagnosis. Diabetologia, 2001;44(2): 156–163.

26. Klein R, Klein BE, Moss SE, Davis MD, DeMets DL. The Wisconsin Epidemiologic Study of Diabetic Retinopathy: IX. Four-year incidence and progression of diabetic retinopathy when age at diagnosis is less than 30 years. Archives of Ophthalmology, 1989;107(2): 237–243.

27. Klein R, Klein BE, Moss SE, Cruickshanks KJ. The Wisconsin Epidemiologic Study of diabetic retinopathy: XIV. Ten-year incidence and progression of diabetic retinopathy. Archives of Ophthalmology, 1994;112(9): 1217–1228.

28. Karamanos B, Porta M, Songini M, Metelko Z, Kerenyi Z, Tamas G, et al. Different risk factors of microangiopathy in patients with type I diabetes mellitus of short versus long duration. The EURODIAB IDDM Complications Study. Diabetologia, 2000;43(3): 348–355.

29. Estacio RO, Jeffers BW, Gifford N, Schrier RW. Effect of blood pressure control on diabetic microvascular complications in patients with hypertension and type 2 diabetes. Diabetes care, 2000;23, B54.

30. Klein R, Klein BE, Moss SE, Cruickshanks KJ. The Wisconsin epidemiologic study of diabetic retinopathy: XVII: The 14-year incidence and progression of diabetic retinopathy and associated risk factors in type 1 diabetes11Proprietary interest: none. Ophthalmology, 1998;105(10): 1801–1815.

31. Chaturvedi N, Sjolie A-K, Stephenson JM, Abrahamian H, Keipes M, Castellarin A, et al. Effect of lisinopril on progression of retinopathy in normotensive people with type 1 diabetes. The Lancet, 1998;351(9095): 28–31.

32. Klein BE, Moss SE, Klein R, Surawicz TS. The Wisconsin Epidemiologic Study of Diabetic Retinopathy: XIII. Relationship of serum cholesterol to retinopathy and hard exudate. Ophthalmology, 1991;98(8): 1261–1265.

33. Wong TY, Cheung N, Tay WT, Wang JJ, Aung T, Saw SM, et al. Prevalence and risk factors for diabetic retinopathy: the Singapore Malay Eye Study. Ophthalmology, 2008;115(11): 1869–1875.

34. Cheung N, Wong TY. Obesity and eye diseases. Survey of ophthalmology, 2007;52(2): 180–195.

35. Klein R, Klein BE, Moss SE, Davis MD, DeMets DL. The Wisconsin Epidemiologic Study of Diabetic Retinopathy: III. Prevalence and risk of diabetic retinopathy when age at diagnosis is 30 or more years. Archives of ophthalmology, 1984;102(4): 527–532.

36. El Haddad OA, Saad MK. Prevalence and risk factors for diabetic retinopathy among Omani diabetics. British journal of ophthalmology, 1998;82(8): 901–906.

37. Do DV, Wang X, Vedula SS, Marrone M, Sleilati G, Hawkins BS, et al. Blood pressure control for diabetic retinopathy. Sao Paulo Medical Journal, 2015;133(3): 278–279.

38. Cichosz SL, Johansen MD, Knudsen ST, Hansen TK, Hejlesen O. A classification model for predicting eye disease in newly diagnosed people with type 2 diabetes. Diabetes research and clinical practice, 2015;108(2): 210–215.

39. Mehlsen J, Erlandsen M, Poulsen PL, Bek T. Individualized optimization of the screening interval for diabetic retinopathy: a new model. Acta ophthalmologica, 2012;90(2): 109–114.

40. Aspelund T, Þórisdóttir Ó, Olafsdottir E, Gudmundsdottir A, Einarsdóttir A, Mehlsen J, et al. Individual risk assessment and information technology to optimise screening frequency for diabetic retinopathy. Diabetologia, 2011;54(10): 2525–2532.

41. Lund S, Aspelund T, Kirby P, Russell G, Einarsson S, Palsson O, et al. Individualised risk assessment for diabetic retinopathy and optimisation of screening intervals: a scientific approach to reducing healthcare costs. British Journal of Ophthalmology, 2015; bjophthalmol–2015.

42. Scanlon PH, Aldington SJ, Leal J, Luengo-Fernandez R, Oke J, Sivaprasad S, et al. Development of a cost-effectiveness model for optimisation of the screening interval in diabetic retinopathy screening. 2015;

43. Ferris F. Early photocoagulation in patients with either type I or type II diabetes. Transactions of the American Ophthalmological Society, 1996;94, 505.

44. Group DRSR et al. Photocoagulation treatment of proliferative diabetic retinopathy: clinical application of Diabetic Retinopathy Study (DRS) findings, DRS Report Number 8. Ophthalmology, 1981;88(7): 583–600.
45. Wharton H, Gibson J, Dodson P. How accurate are photographic surrogate markers used to detect macular oedema in the English national screening programme? European Journal of Ophthalmology, 2011;21(3):