UCLA

UCLA Previously Published Works

Title

DRRisk: A Web-based tool to Assess the Risk of Diabetic Retinopathy through Machine Learning on Electronic Health Records.

Permalink

https://escholarship.org/uc/item/3dj8z8nf

Authors

Gandhi, Meghal
Daskivich, Lauren
Ogunyemi, Omolola

Publication Date

2022

Peer reviewed

DRRisk: A Web-based tool to Assess the Risk of Diabetic Retinopathy through Machine Learning on Electronic Health Records

Meghal Gandhi, MS[1,2], Lauren Patty Daskivich, MD, MSHS[2,4], Omolola I. Ogunyemi, PhD[1,2,3],

[1]Center for Biomedical Informatics, [2]Charles R. Drew University of Medicine and Science, Los Angeles, CA; [3]University of California Los Angeles, Los Angeles, CA; [4]Los Angeles County Department of Health Services, Los Angeles, CA;

Abstract

Objective: We developed a web-based tool for diabetic retinopathy (DR) risk assessment called DRRisk (https://drandml.cdrewu.edu/) using machine learning on electronic health record (EHR) data, with a goal of preventing vision loss in persons with diabetes, especially in underserved settings.

Methods: DRRisk uses Python's Flask framework. Its user-interface is implemented using HTML, CSS and Javascript. Clinical experts were consulted on the tool's design.

Results: DRRisk assesses current DR risk for people with diabetes, categorizing their risk level as low, moderate, or high, depending on the percentage of DR risk assigned by the underlying machine learning model.

Discussion: A goal of our tool is to help providers prioritize patients at high risk for DR in order to prevent blindness.

Conclusion: Our tool uses DR risk factors from EHR data to calculate a diabetic person's current DR risk. It may be useful for identifying unscreened diabetic patients who have undiagnosed DR.

Introduction

According to recent statistics published by the Centers for Disease Control and Prevention (CDC), 37.3 million people have diabetes in the United States.[1] Diabetic retinopathy (DR) is a complication of diabetes that can cause vision loss and/or blindness. DR is the leading cause of blindness in US adults. If caught early, it is treatable. In medically underserved settings, teleretinal diabetic retinopathy screening in primary care clinics has been shown to be an acceptable standard of care that allows safety net health providers to meet the needs of a large volume of diabetic patients when a limited number of eye specialists are available for in person examinations. [2,3,4]

To help identify diabetic persons with undiagnosed DR in safety net settings, we developed a machine learning (ML) model trained using clinical data from patients in the Los Angeles County Department of Health Services healthcare system[5] and then created an online tool called DRRisk based on that ML model. The goal of this online tool is to help clinicians in medically underserved and under-resourced settings identify patients at high risk for DR and ensure they receive appropriate guideline-concordant eye care in a timely manner. Patients who are assessed as having a high risk of DR by this online tool and who have also missed their annual diabetic retinopathy screening examination can be contacted to schedule a teleretinal screening or an in-person eye examination with an eye care provider for DR.

Annual screening for diabetic retinopathy is recommended by the American Diabetes Association for diabetic patients.[6] Medically underserved and under resourced areas struggle with a shortage of eye specialists, which makes it difficult for safety net providers to meet the recommendation of the American Diabetes Association. In such cases, our online tool may be an effective means to assess patients' DR risk and conduct outreach in a timely manner. In addition, DRRisk may be helpful to categorize patients' severity of DR, so that clinicians can prioritize those who are at high risk and refer them for examination and treatment as early as possible.

DRRisk generates a person's DR assessment using risk factor data (e.g., duration of diabetes, Hemoglobin A1c levels, etc.) obtained from electronic health records. Our application (https://drandml.cdrewu.edu/) allows users to input both clinical and demographic data and generates a DR risk assessment based on the data input. The purpose of this article is to describe the motivation behind DRRisk, to explain the process behind its assessment of risk, and to provide an overview of its features.

Methods

We created a web-based DR risk assessment tool, DRRisk, based on an underlying ML model that provides a risk probability for DR. We selected the name DRRisk to reflect our goal of assessing a diabetic person's current risk of having diabetic retinopathy. The central component of this tool is a machine-learning model developed using the electronic health records of patients with type 1 and type 2 diabetes seen between January 1, 2015 to December 31, 2017 at LACDHS facilities.[5] In this section, we describe the web-based tool's design and implementation, including the choices made to preserve utility for clinicians as well as user-friendliness. Figure 1 gives an overview of the DR Risk tool's assessment workflow.

Figure 1. DRRisk DR assessment flow

As shown in figure 1, the tool asks the user to provide values for the input predictors on the predict page. Out of 14 predictors, 6 predictors are mandatory in order to get the DR assessment. These 6 risk factors are Duration of Diabetes (Years), Hemoglobin A1c, Blood Urea Nitrogen (BUN), Sex, Ethnicity and Insulin Dependence.

Once the user enters the required data and clicks on *Assess DR Risk,* the tool sends the data to the server for further processing and redirects users to the result page. There are three main steps performed to get the data ready for the ML model: (1) getting imputed data for any missing predictors by using K-Nearest Neighbor (KNN) imputation based on the training data for the ML model; (2) performing data standardization and normalization for numerical predictors; (3) performing One Hot Encoding on categorical predictors using one hot encoding training data object. Once these steps are completed, the ML model will execute the risk assessment and generate the percentage risk as a result. The result page shows the percentage risk, a pie chart of the assessment analysis, and predicted DR risk category as low, moderate or high risk. The tool will categorize the risk in percentages as: (1) patient with low risk DR (i.e., less than 25% risk), (2) patient with moderate risk DR (i.e., 25% to 55% risk) and (3) patient with high risk DR (i.e., greater than 55% risk). In situations where incorrect data is inputted or the data has changed, the user can update previously entered data and regenerate the DR risk assessment.
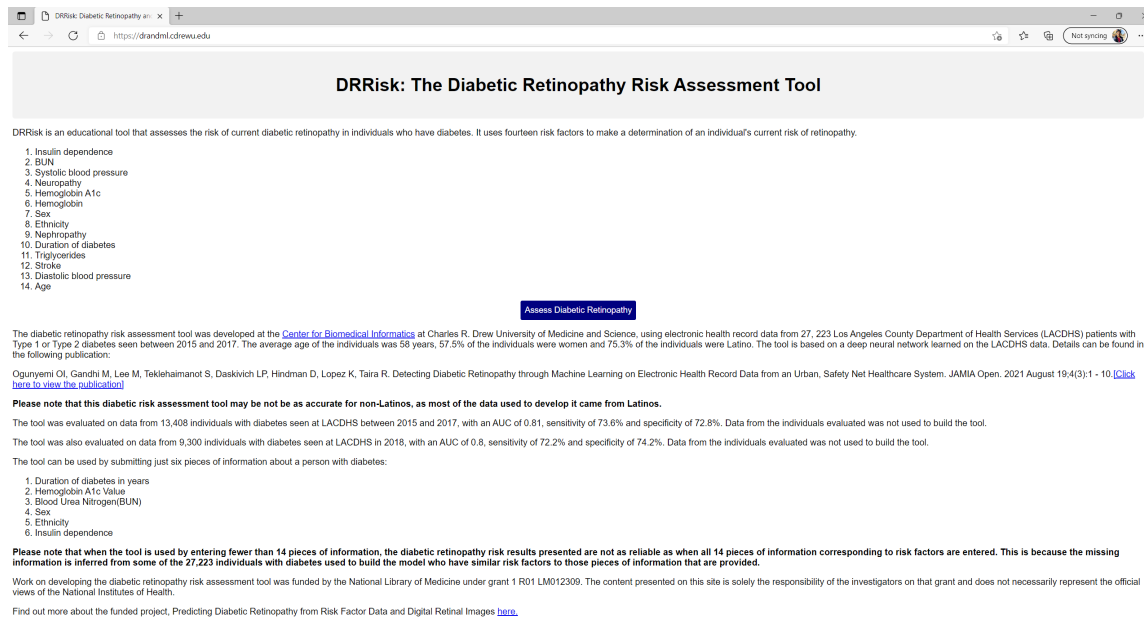
Figure 2. DRRisk tool introduction page

Users start the DR assessment by browsing our tool's website. Figure 2 is a screen shot of the DRRisk landing page, found at https://drandml.cdrewu.edu/. It describes the risk factors used in this tool along with the published study details. Study details include the external validation results along with the data sources used to develop the machine-learning model. Users are required to click on the *Assess Diabetic Retinopathy* button in order to perform the DR risk assessment.

Web-based frameworks assessed and web design

To develop the web-based tool for DR risk assessment, we considered using either the R-Shiny web framework backed by R programming or the Python flask web framework backed by Python. We compared various aspects of each framework in table 1 to finalize this decision.

Table 1: Comparison between Python Flask and R-Shiny framework

ID	R-Shiny	Python Flask
1	It is most suitable for building data-dashboard like applications.	It is most suitable for building web applications.
2	No prior knowledge of HTML, CSS and/or JavaScript is required.	It requires HTML, CSS, and/or JavaScript experience along with Python experience.
3	It is great for users who have zero front-end development experience.	It is great for developers with front-end experience. In Python Flask, the Python code is completely decoupled from the web languages.
4	R-Shiny apps work best with a Linux based server.	Python Flask apps can be hosted to both Linux and Windows based servers.

For the web app deployment environment, we used a Microsoft Azure-based Windows cloud virtual machine, which is equipped with a Windows server Internet Information Services (IIS). R-shiny does not have any official documentation available to deploy web apps in Windows IIS. Alternatively, Microsoft has provided a well-defined deployment process for Python web apps using common gateway interface.[7] That led us to choose the Python Flask framework for our DRRisk tool development.

The components of DRRisk include a Microsoft Azure cloud-based platform running IIS (Internet Information Services) server and Python Flask based web framework. The most essential feature of our tool is to assess a person's diabetic retinopathy risk using risk factor data from a patient's medical records. In case of incorrect inputs, this tool allows users to change/update inputted data easily after getting an assessment result.

Consultations with experts to determine minimum number of features that must be input

DRRisk's risk factor data elements were selected using a feature subset selection process in the machine-learning model development phase of our study.[5] These 14 risk factors are Duration of Diabetes, Hemoglobin A1C, Blood Urea Nitrogen (BUN), Age, Systolic Blood Pressure, Diastolic Blood Pressure, Hemoglobin, Sex, Ethnicity, Insulin Dependence, Nephropathy, Neuropathy, Stroke, and Triglycerides. We consulted an endocrinologist, a primary care doctor and an ophthalmologist on the ideal set of risk factors from these original fourteen that would be required for the model, because there was a concern that fourteen variables would be too many for busy clinicians to input. As an outcome of this consultation, six risk factors were selected out of fourteen based on a balance between their ease of acquisition and their importance for accuracy of the model. These six risk factors are divided into three numerical (Duration of Diabetes, Hemoglobin A1C, and Blood Urea Nitrogen) and three categorical (Sex, Ethnicity, and Insulin Dependence) risk factors.

Imputation approach when fewer than fourteen risk factors are input

Our tool's machine learning model requires all fourteen risk factors in order to generate a DR risk assessment. If users enter only six risk factors, then the values for the remaining eight will be imputed using a KNN imputer object on the backend (KNN imputation for DRRisk is based on the training data for the ML model).[5] Similarly, if users enter ten risk factors then the remaining four will be imputed. We encourage users to provide as many risk factors as possible to get the most accurate DR risk assessment.

Determination of risk categories (low, moderate and high) and cut-offs

To decide on the risk cut-offs for each of the three risk categories (low, moderate, and high) we consulted with domain experts in ophthalmology, internal medicine, and endocrinology. For ease of interpretation in our web-based, end-user facing tool, we converted this risk probability into three categories: low (less than 25% risk of DR), moderate (25% to 55% risk of DR), and high (greater than 55% risk of DR).

Validation and range checking for user input

We coded custom validations for all the input risk factors using JavaScript/CSS and HTML. One of the custom validations is to force users to provide the mandatory six risk factors marked with a red asterisk. These six risk factors are the minimum required for the tool to produce a useful assessment, otherwise it will rely more on imputed data which will ultimately decrease the accuracy of the assessment.

Our study included data for persons with diabetes who are 18 years of age or older. We also included validations in our tool to ensure that patient age is greater than the duration of diabetes. For example, if a patient's age is 20 years and duration of diabetes is entered as 25 years, the tool will show an error message.

To determine the plausible ranges for each numerical risk factor, we created a codebook based on appropriate values from the biomedical literature. When there was no established plausible range for a risk factor, we met with an ophthalmologist (LPD) to determine acceptable ranges based on data from a safety net healthcare setting. We then coded the minimum and maximum acceptable values, to provide validation for each numerical risk factor in our tool. This prevents users from entering biologically implausible values for the risk factors that have numerical values. Table 2 lists the minimum and maximum values for each of the risk factors that are numerical.

Table 2: Minimum and maximum value for numerical risk factors – custom validations

Input Risk Factor	Min. value used in our tool	Max. value used in our tool
Duration of Diabetes(Years)	0 year(s)	120 year(s)
Hemoglobin A1C	3.7 %	20 %
Blood Urea Nitrogen(BUN)	1 mg/dL	100 mg/dL
Age	18 years	120 years
Systolic Blood Pressure	80 mmHg	220 mmHg
Diastolic Blood Pressure	40 mmHg	130 mmHg
Hemoglobin	0.6 g/dL	25 g/dL
Triglycerides	1 mg/dL	500 mg/dL

Machine learning model for DRRisk

Our DRRisk tool uses a Deep Neural Network model for risk assessment, the ML model (out of five evaluated) that had the best AUC results on the test set and the external validation set in our published study.[5]

Results

We show in this section the impact of our tool with examples of patients having low, moderate, and high risk for DR in figures 3, 4 and 5 respectively.
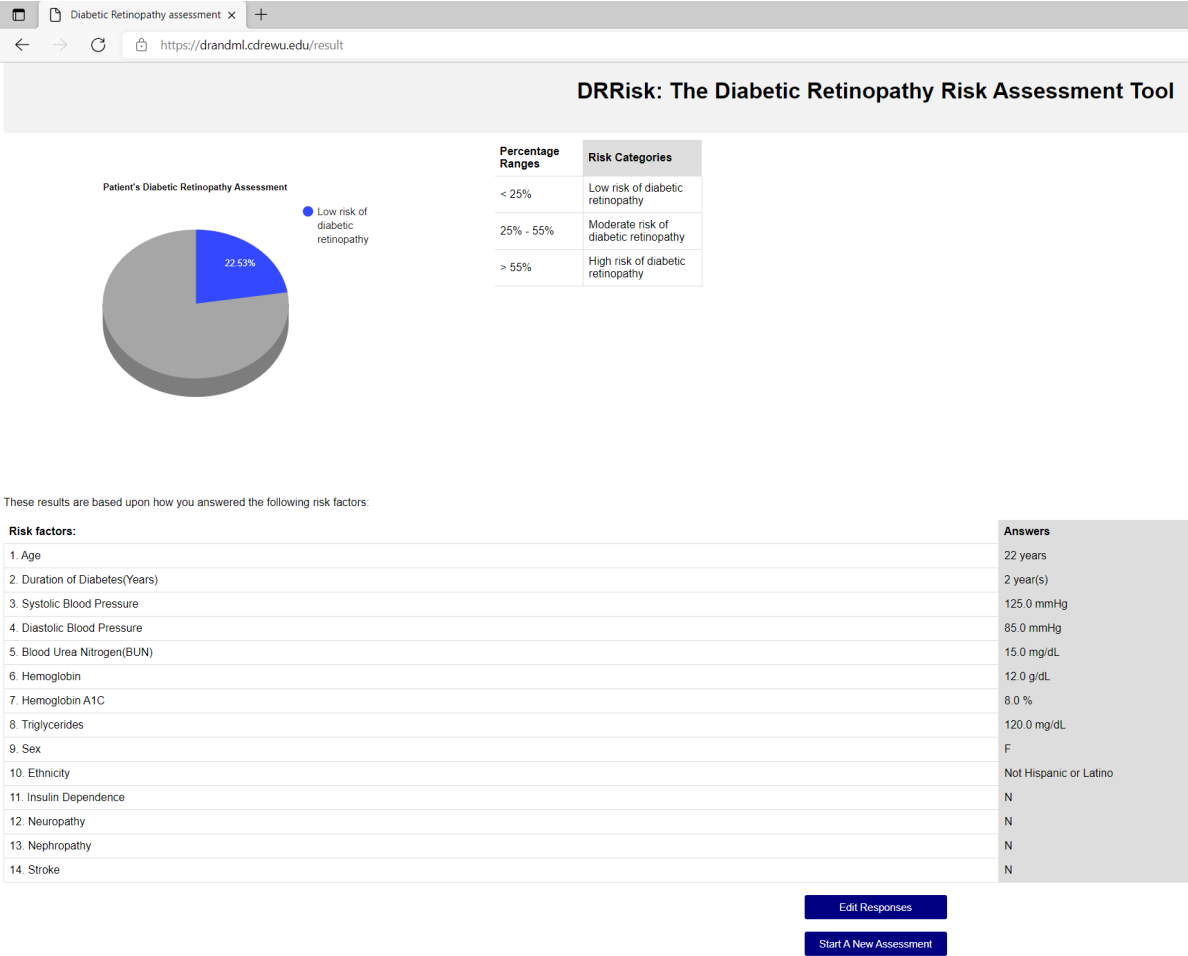

Figure 3. Patient with low risk DR

Figure 3 shows an example of a patient who has a low risk of DR according to DRRisk's assessment. Table 3 shows all the risk factor values for the low-risk DR patient.
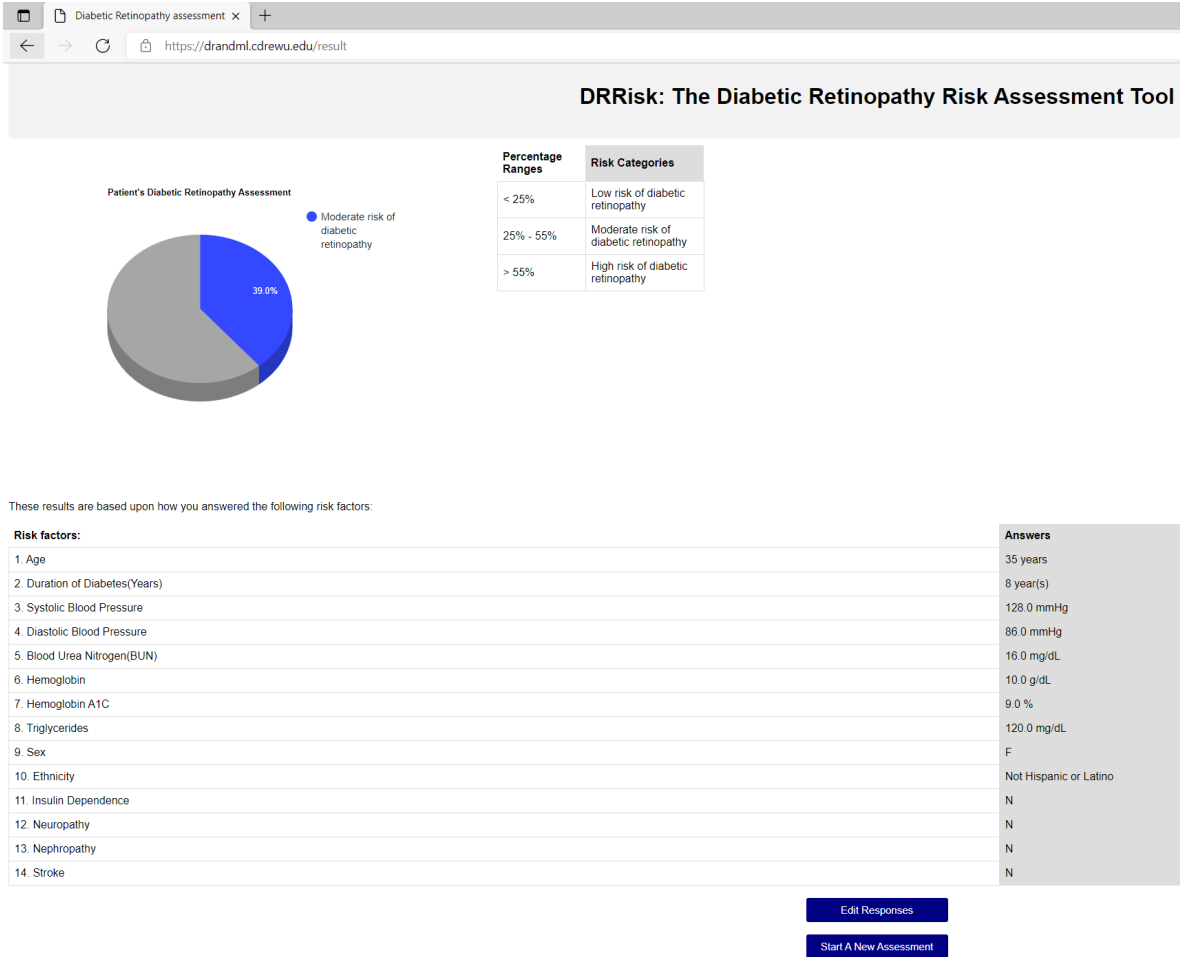
DRRisk: The Diabetic Retinopathy Risk Assessment Tool

Patient's Diabetic Retinopathy Assessment

● Moderate risk of diabetic retinopathy

39.0%

Percentage Ranges	Risk Categories
< 25%	Low risk of diabetic retinopathy
25% - 55%	Moderate risk of diabetic retinopathy
> 55%	High risk of diabetic retinopathy

These results are based upon how you answered the following risk factors:

Risk factors:	Answers
1. Age	35 years
2. Duration of Diabetes(Years)	8 year(s)
3. Systolic Blood Pressure	128.0 mmHg
4. Diastolic Blood Pressure	86.0 mmHg
5. Blood Urea Nitrogen(BUN)	16.0 mg/dL
6. Hemoglobin	10.0 g/dL
7. Hemoglobin A1C	9.0 %
8. Triglycerides	120.0 mg/dL
9. Sex	F
10. Ethnicity	Not Hispanic or Latino
11. Insulin Dependence	N
12. Neuropathy	N
13. Nephropathy	N
14. Stroke	N

Edit Responses

Start A New Assessment

Figure 4: Patient with moderate risk DR

Figure 4 shows an example of a patient who has moderate risk of DR according to DRRisk's assessment. Table 3 shows all the risk factor values for the moderate risk DR patient.
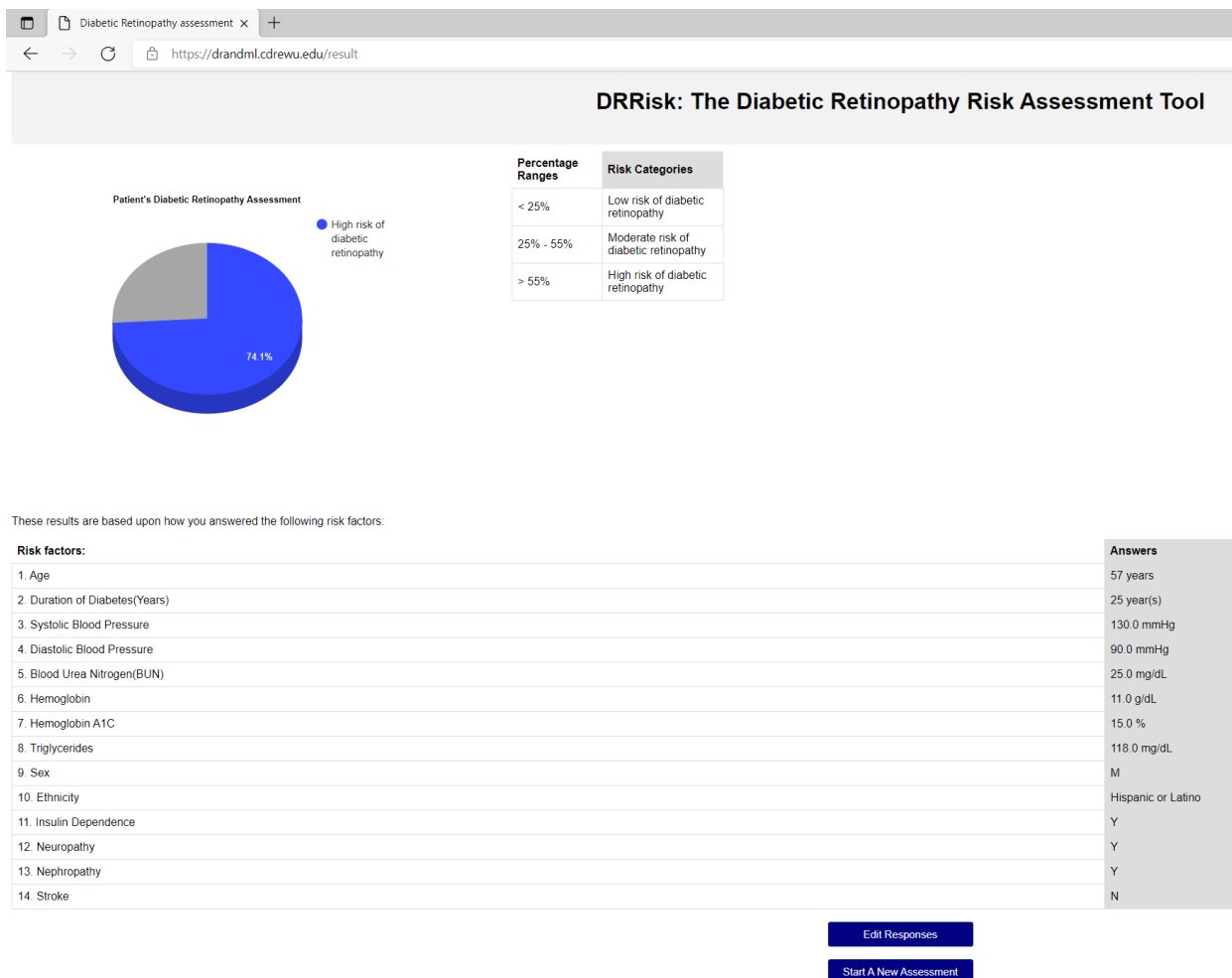
Figure 5: Patient with high risk DR

Figure 5 shows an example of a patient who has high risk of DR according to DRRisk's assessment. Table 3 shows all the risk factor values for the high-risk DR patient.

Table 3. Example of input risk factor values – low, moderate and high risk DR patients

Input Risk Factor	Low Risk Values	Moderate Risk Values	High Risk Values
Age	22 years	35 years	57 years
Duration of Diabetes(Years)	2 year(s)	8 year(s)	25 year(s)
Systolic Blood Pressure	125.0 mmHg	128.0 mmHg	130.0 mmHg
Diastolic Blood Pressure	85.0 mmHg	86.0 mmHg	90.0 mmHg
Blood Urea Nitrogen(BUN)	15.0 mg/dL	16.0 mg/dL	25.0 mg/dL
Hemoglobin	12.0 g/dL	10.0 g/dL	11.0 g/dL
Hemoglobin A1C	8.0 %	9.0 %	15.0 %
Triglycerides	120.0 mg/dL	120.0 mg/dL	118.0 mg/dL
Sex	F	F	M
Ethnicity	Not Hispanic or Latino	Not Hispanic or Latino	Hispanic or Latino
Insulin Dependence	N	N	Y
Neuropathy	N	N	Y
Nephropathy	N	N	Y
Stroke	N	N	N

458

Discussion

With the goal of providing an estimate of current DR risk for a person with diabetes, using risk factor data entered by either a clinician or patients themselves, we created a user-friendly web-based tool called DRRisk.

We compared our DRRisk tool with similar ones currently in use or recently developed for diabetic retinopathy risk assessment. One such tool is running on the American Diabetes Association website, generating a patient's one-year risk of developing DR.[8] It was built based on data from an Icelandic study and like ours, is focused on patients who have Type-1 or Type-2 diabetes. Another DR risk assessment tool was developed as part of diabetic retinopathy research funded by Google.[9] The Google study used digital retinal images of diabetic patients collected through teleretinal DR screening initiatives to predict a patient's risk of developing diabetic retinopathy within two years. Both of the above tools predict a person's future DR risk.

The major difference between our DRRisk tool and the above tools is that DRRisk focuses on current risk, not future risk. Additionally, our DRRisk tool uses clinical and demographic data from electronic health records (EHRs) to generate DR risk assessment rather than digital retinal images. Our study focuses on underserved and under-resourced settings where the capacity for annual DR screening is limited. Since our tool requires only EHR data, it could be a useful asset for clinicians in these settings looking to target limited resources to those at highest risk for diabetic eye disease.

In terms of future enhancements to DRRisk, we will add batch mode processing of Fast Healthcare Interoperability Resources (FHIR) data. We have begun work on this, and once this feature is fully integrated, our tool will be able to assess current DR risk for up to 200 patients at a time in batch mode and generate an assessment report in pdf format. We will update our ML model with training data from additional medically underserved settings to broaden the racial and ethnic data that underpin the model and will integrate these changes in our tool to make it more generalizable. We will also expand our research to include an evaluation of variable importance for an individual's DR assessment in our tool.

Conclusion

DRRisk is a web-based tool that leverages the power of machine learning to provide end users with current DR risk, with a goal of improving access to care for diabetic retinopathy in under-resourced settings. It is geared towards patients in underserved populations, who often do not receive recommended annual eye examinations for diabetic retinopathy, and their healthcare providers. Potential benefits of a tool like DRRisk are to make patients aware of their current DR risk and to help clinicians in prioritizing outreach to patients at high-risk for DR to decrease their risk of vision loss and potential blindness from DR.

Acknowledgement

References

1. National Diabetes Statistics Report. Centers for Disease Control and Prevention. https://www.cdc.gov/diabetes/data/statistics-report/index.html. Accessed January 18, 2022.
2. Daskivich, L. P., Vasquez, C., Martinez, C., Jr, Tseng, C. H., & Mangione, C. M. (2017). Implementation and Evaluation of a Large-Scale Teleretinal Diabetic Retinopathy Screening Program in the Los Angeles County Department of Health Services. *JAMA internal medicine*, *177*(5), 642–649. https://doi.org/10.1001/jamainternmed.2017.0204
3. Mansberger, S. L., Sheppler, C., Barker, G., Gardiner, S. K., Demirel, S., Wooten, K., & Becker, T. M. (2015). Long-term Comparative Effectiveness of Telemedicine in Providing Diabetic Retinopathy Screening Examinations: A Randomized Clinical Trial. *JAMA ophthalmology*, *133*(5), 518–525. https://doi.org/10.1001/jamaophthalmol.2015.1

4. Mehraban Far P, Tai F, Ogunbameru A*, et al.* Diagnostic accuracy of teleretinal screening for detection of diabetic retinopathy and age-related macular degeneration: a systematic review and meta-analysis *BMJ Open Ophthalmology* 2022;**7:**e000915. doi: 10.1136/bmjophth-2021-000915

5. Ogunyemi OI, Gandhi M, Lee M, et al. Detecting diabetic retinopathy through machine learning on electronic health record data from an urban, safety net healthcare system. *JAMIA Open*. 2021;4(3). doi:10.1093/jamiaopen/ooab066

6. Solomon SD, Chew E, Duh EJ, Sobrin L, Sun JK, VanderBeek BL, et al. Diabetic Retinopathy: A Position Statement by the American Diabetes Association. Diabetes Care. 2017 Mar;40(3):412-8.

7. Configure python web apps for IIS - Visual Studio (windows). Visual Studio (Windows) | Microsoft Docs. https://docs.microsoft.com/en-us/visualstudio/python/configure-web-apps-for-iis-windows?view=vs-2022

8. Diabetic retinopathy risk test. Diabetic Retinopathy Risk Test | ADA. https://diabetes.org/diabetes/eye-health/retinopathy-risk. Accessed February 28, 2022.

9. Bora A., Balasubramanian S., Babenko B., Virmani S., Venugopalan S., Mitani A., de Oliveira Marinho G., Cuadros J., Ruamviboonsuk P., Corrado G. S., Peng L., Webster D. R., Varadarajan A. V., Hammel N., Liu Y., & Bavishi P. (2021). Predicting the risk of developing diabetic retinopathy using Deep Learning. *The Lancet Digital Health*, *3*(1). https://doi.org/10.1016/s2589-7500(20)30250-8.