A Novel Hyperparameter Search Approach for Accuracy and Simplicity in Disease Prediction Risk Scoring

Yajun Lu

Assistant Professor of Analytics & Operations Management
Jacksonville State University
ylu@jsu.edu

Joint work with:

Thanh Duong[a], Zhuqi Miao[b], Thanh Thieu[a], Jivan Lamichhane[c], Abdulaziz Ahmed[d], and Dursun Delen[e]

[a]Moffitt Cancer Center, [b]State University of New York at New Paltz
[c]New York Upstate Medical University, [d]University of Alabama at Birmingham, [e] Oklahoma State University

October 21, 2024

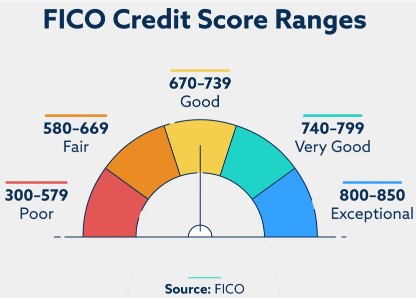
Outline

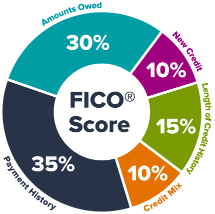
Outline

Credit Scoring: A Win-Win for Lenders & Borrowers

1. Credit scores serve as a valuable tool that benefits credit providers by minimizing risk and streamlining lending processes.

2. It empowers consumers with financial opportunities, better loan terms, and a pathway to financial education and responsibility.





Are there similar risk scores in healthcare?

A Risk Score Example: LACE Score (Van Walraven et al., 2010)

Component	Value	Score
L: Length of Stay (days)	< 1	0
	1	1
	2	2
	3	3
	4-6	4
	7-13	5
	≥ 14	7
A: Acuity (Emergent) Admission	No	0
	Yes	3
C: Charlson Comorbidity Index Score	0	0
	1	1
	2	2
	3	3
	≥ 4	5
E: Emergency Department Visits During Previous 6 Months	1	1
	2	2
	3	3
	≥ 4	4

Risk Scores in Healthcare

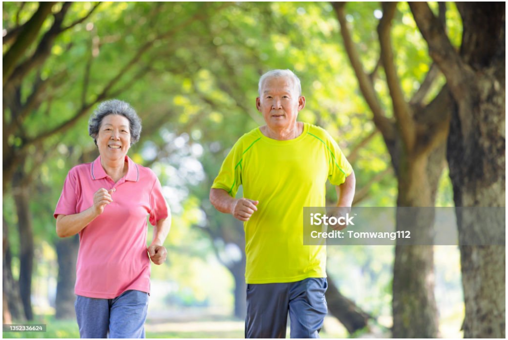
Image: dashboardmd.com

- LACE (Van Walraven et al., 2010) and HOSPITAL (Donzé et al., 2013) for anticipating death or readmission after hospital discharge
- Framingham Risk Scores for predicting coronary heart disease (Wilson et al., 1998; DAgostino Sr et al., 2008)
- IScore (Saposnik et al., 2011) for predicting death and disability after an acute stroke
- Mortality Risk Score (Austin and van Walraven, 2011) for estimating mortality in adults

Benefits of Risk Scores

Image: Hospital Pharmacy Europe

Enhancing patient risk stratification for efficient medical resource allocation



Empower patients for positive health behavior change

Steps in Developing Risk Scoring Systems

The main idea is to map regression units to integer points by approximating the linear combination $\sum_{i=1}^{n} \beta_i X_i$ in the logistic regression model (Sullivan et al., 2004):

$$P(Y=1 \mid X) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_n X_n)}}$$

Variable	β	Level	Reference Value	Height	Risk	Score
Sodium	0.092	< 136	$131.5^{\dagger}$	0	0	0
		136 –144	140	8.5	0.78	1
		≥ 144	146.5	15	1.38	2

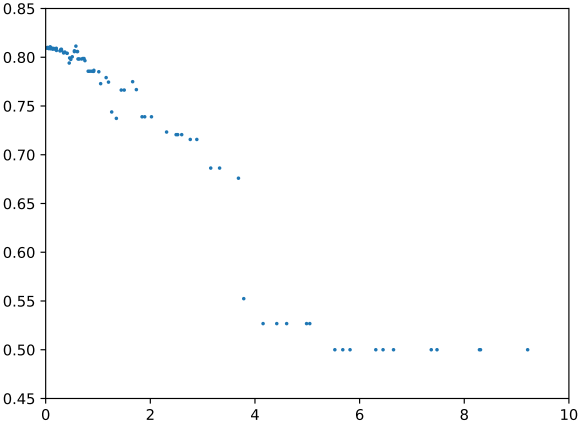
1. Meaningful cutoffs $\longrightarrow$ "Heights"
2. Map "height" to risk
   - Risks are $\beta_{Sodium} \times 0$, $\beta_{Sodium} \times 8.5$, $\beta_{Sodium} \times 15$
3. Map risk to score
   - Set the constant B: number of regression units that reflect 1 point in the scoring system
   - All risks divided by B, then round to integers

Impact of B Selection on Risk Score Performance & Interpretability

Currently, there are no specific rules for selecting B:

B	AUC	Score Range
Large	Low	Small
Small	High	Large[a]

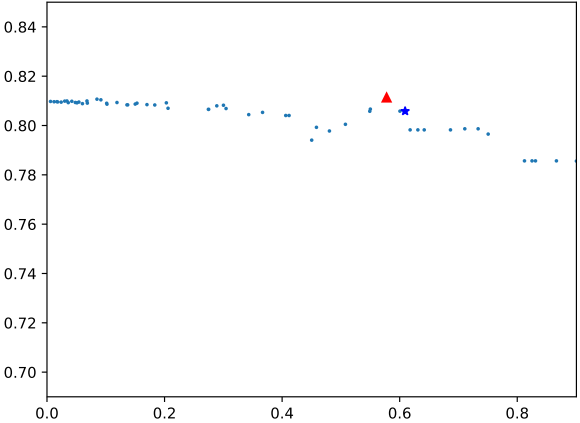
[a] When $B = B_{\min}$, the score range is 0–2,510



A relationship between constant B and area under the receiver operating characteristic curve (AUC) for predicting Diabetic Retinopathy.

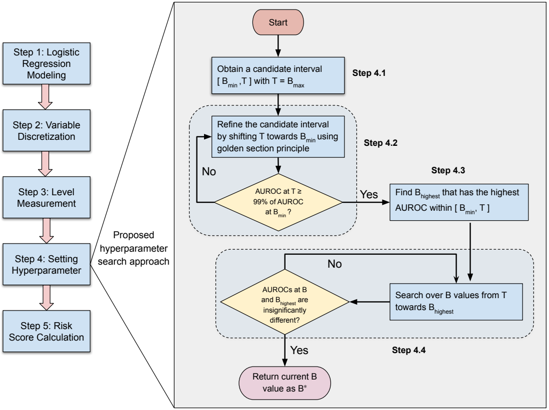
A Close-up View of *B* vs AUC

- How can we select a *B* as large as possible while maintaining a reasonably high AUC value?

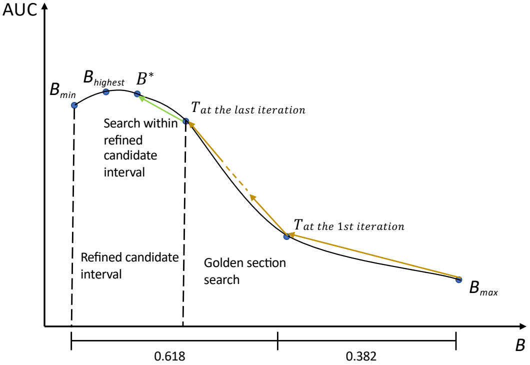


Outline

Risk Score Derivation Framework



Flowchart of the risk scoring system and our refinements in Step 4



An illustration of our search approach

Outline

Case Study on Diabetic Retinopathy (DR) Prediction

Normal Vision

Diabetic Retinopathy Vision

Annual eye examination is so far critical for early screening of DR. Unfortunately the adherence is poor

- DR is a vision-threatening microvascular complication of diabetes (Yau et al., 2012).
- DR is symptomless in early stages
- Vision loss can be halted but not reversed

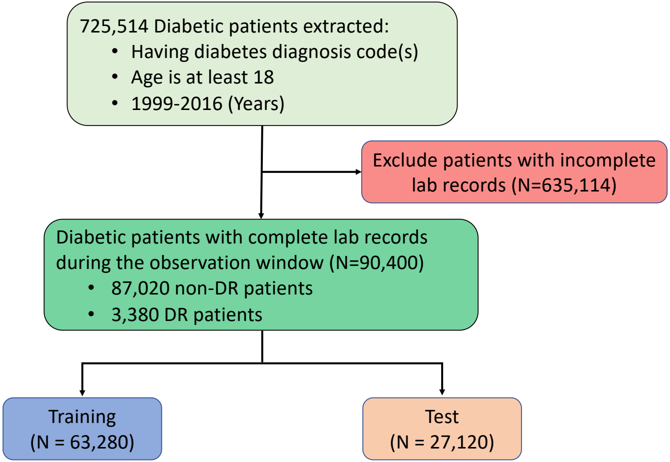
Current solution:

- Early detection: Annual eye examination guideline
- Timely treatment: Laser surgery
- Appropriate follow-up

Barriers to early detection:

- Approximately 43% adherence rate to annual eye examinations (Fisher et al., 2016)
- A lack of comprehensive screening programs in rural areas

Data Preparation



The workflow for the cohort development

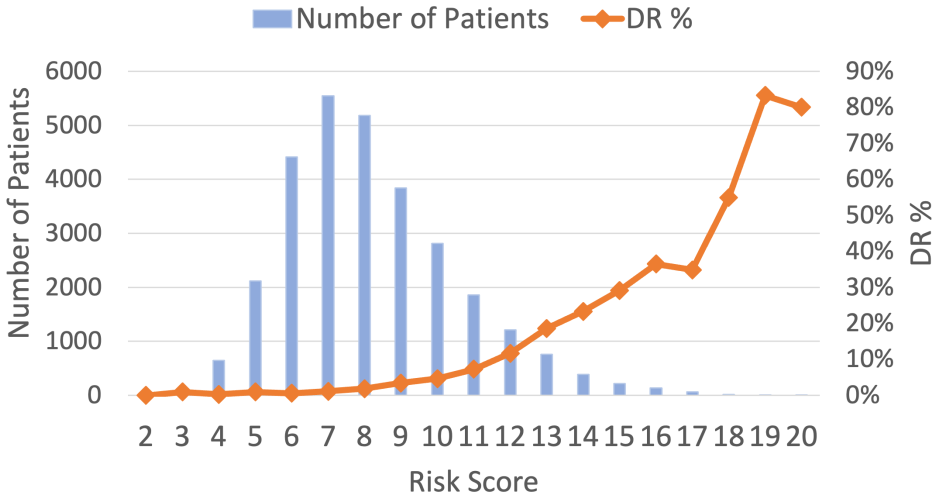
- Oracle Cerner Health Facts® EHR data
- 31 Variables:
  - Demographics: gender, race, and age
  - Complications: nephropathy and neuropathy
  - Duration of diabetes
  - 25 different routine blood tests

Risk Score for DR

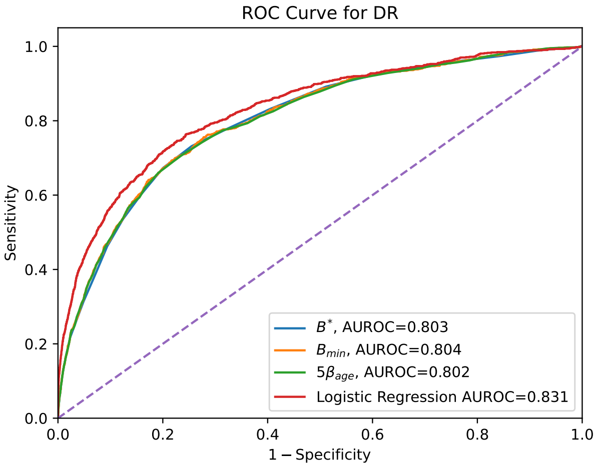
Variable	Levels	Risk Score for DR		
		B^*	$5\beta_{age}$	$B_{\min}$
Neuropathy	No	0	0	0
	Yes	3	11	638
Nephropathy	No	0	0	0
	Yes	2	6	365
Creatinine	<0.5	0	0	0
	[0.5, 1)	1	4	203
	[1, 1.5)	2	9	502
	[1.5, 2)	4	14	801
	≥ 2	7	24	1,357
HbA1c	<6	0	0	0
	[6, 8)	2	7	383
	[8, 10)	4	13	767
	[10, 12)	6	20	1,150
	≥ 12	8	30	1,725
Diabetes Duration	<1	0	0	0
	[1, 2)	0	2	96
	[2, 3)	1	3	192
	[3, 4)	1	5	287
	≥ 4	4	15	833
White Blood Cell	<4	4	14	773
	[4, 6)	3	12	694
	[6, 8)	3	10	589
	[8, 12)	2	8	431
	≥ 12	0	0	0
Glucose	<60	0	0	0
	[60, 80)	0	1	76
	[80, 100)	1	3	166
	[100, 200)	2	8	434
	≥ 200	7	24	1,393

Variable	Levels	Risk Score for DR		
		B^*	$5\beta_{age}$	$B_{\min}$
Age	<35	3	12	704
	[35, 50)	2	9	515
	[50, 65)	2	6	343
	[65, 75)	1	4	200
	[75, 85)	0	2	86
	≥ 85	0	0	0
Hematocrit	<30	6	21	1,215
	[30, 35)	5	16	933
	[35, 40)	4	13	725
	[40, 50)	2	7	415
	≥ 50	0	0	0
Sodium	<136	0	0	0
	[136, 144)	3	11	620
	≥ 144	5	19	1094
BUN	<11	0	0	0
	[11, 15)	0	0	3
	[15, 19)	0	0	7
	[19, 27)	0	0	14
	≥ 27	0	0	24
Anion Gap	<5	3	11	648
	[5, 7)	3	10	575
	[7, 10)	2	8	453
	[10, 12)	2	6	330
	[12, 17)	1	3	159
	≥ 17	0	0	0
Race	African American	1	4	248
	Other	1	2	124
	Caucasian	0	0	0

Risk Score Distributions



Risk Score AUC & Scale Comparisons



	B^*	$5\beta_{age}$	B_{min}
AUC	**0.803**	0.802	0.804
Scale	**0–53**	0–191	0–11,018

Outline

Concluding remarks

- Simple yet accurate risk scores for patient stratification and resource allocation
- User-friendly tools motivating healthier habits, reducing healthcare demand
- Our case studies on diabetic retinopathy (DR) and hip fracture readmission (HFR) demonstrate the effectiveness of our approach in generating simple, accurate risk scores for disease prediction.
- Potential general framework for simpler disease risk scoring

Concluding remarks

- Simple yet accurate risk scores for patient stratification and resource allocation
- User-friendly tools motivating healthier habits, reducing healthcare demand
- Our case studies on diabetic retinopathy (DR) and hip fracture readmission (HFR) demonstrate the effectiveness of our approach in generating simple, accurate risk scores for disease prediction.
- Potential general framework for simpler disease risk scoring

Manuscript

Code

Q & A

ylu@jsu.edu

http://yajunlu.com

Reference I

P. C. Austin and C. van Walraven. The mortality risk score and the adg score: two points-based scoring systems for the johns hopkins aggregated diagnosis groups (adgs) to predict mortality in a general adult population cohort in ontario, canada. *Medical care*, 49(10):940, 2011.

R. B. DAgostino Sr, R. S. Vasan, M. J. Pencina, P. A. Wolf, M. Cobain, J. M. Massaro, and W. B. Kannel. General cardiovascular risk profile for use in primary care: the framingham heart study. *Circulation*, 117 (6):743–753, 2008.

J. Donzé, D. Aujesky, D. Williams, and J. L. Schnipper. Potentially avoidable 30-day hospital readmissions in medical patients: derivation and validation of a prediction model. *JAMA internal medicine*, 173(8): 632–638, 2013.

M. D. Fisher, Y. Rajput, T. Gu, J. R. Singer, A. R. Marshall, S. Ryu, J. Barron, and C. MacLean. Evaluating adherence to dilated eye examination recommendations among patients with diabetes, combined with patient and provider perspectives. *American Health & Drug Benefits*, 9(7):385, 2016.

G. Saposnik, M. K. Kapral, Y. Liu, R. Hall, M. O'Donnell, S. Raptis, J. V. Tu, M. Mamdani, and P. C. Austin. Iscore: a risk score to predict death early after hospitalization for an acute ischemic stroke. *Circulation*, 123(7):739–749, 2011.

L. M. Sullivan, J. M. Massaro, and R. B. DAgostino Sr. Presentation of multivariate data for clinical use: The framingham study risk score functions. *Statistics in medicine*, 23(10):1631–1660, 2004.

Reference II

C. Van Walraven, I. A. Dhalla, C. Bell, E. Etchells, I. G. Stiell, K. Zarnke, P. C. Austin, and A. J. Forster. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. *Canadian Medical Association Journal*, 182(6):551–557, 2010.

P. W. Wilson, R. B. DAgostino, D. Levy, A. M. Belanger, H. Silbershatz, and W. B. Kannel. Prediction of coronary heart disease using risk factor categories. *Circulation*, 97(18):1837–1847, 1998.

J. W. Yau, S. L. Rogers, R. Kawasaki, E. L. Lamoureux, J. W. Kowalski, T. Bek, S.-J. Chen, J. M. Dekker, A. Fletcher, J. Grauslund, et al. Global prevalence and major risk factors of diabetic retinopathy. *Diabetes care*, 35(3):556–564, 2012.

Backup slides

Data Preparation for DR Prediction

- Idea: Use data before the first DR diagnosis (Ng et al. 2016)
  - Time windows
  - Variable aggregation