UCLA-CDU CFAR Clinical Science Core

Clinical Discovery Seminar

Presented in Collaboration with the

Artificial Intelligence Methods Scientific Working Group (AIM-SWG)

AIM-SWG

Using Electronic Health Record Data for Machine Learning to Predict Risk in the Healthcare Safety Net

Presented By

Omolola Ogunyemi, PhD, FACMI

Director, Center for Biomedical Informatics

Professor & Chair, Dept. of Preventive and Social Medicine

Charles R. Drew University of Medicine and Science

UCLA CDU CFAR

CENTER FOR AIDS RESEARCH

Clinical Discovery Seminar Objectives

- Create a community of clinical, behavioral, and translational HIV researchers at UCLA and CDU and beyond
- Share innovative work by CFAR investigators utilizing NIH and other funded data and sample repositories
- Introduce investigators to opportunities to propose new work/projects using existing data and sample repositories
- Stimulate new ideas and collaboration
- Support investigators in proposing analyses to these cohorts

Clinical Science Core Events

- To access past seminars/workshops and learn about future events, please visit our website: **https://cfar.ucla-cdu.org/cores/clinical-science-core/**

Core Contact Information:
Core Director – Kara Chew, MD, MS
Core Program Manager – Stephanie Buchbinder, MPH
Core email: cfarclinicalcore@mednet.ucla.edu

Using Electronic Heath Record Data for Machine Learning to Predict Risk in the Healthcare Safety Net

Omolola Ogunyemi, PhD, FACMI
Director, Center for Biomedical Informatics
Professor & Chair, Dept. of Preventive and Social Medicine
Charles R. Drew University of Medicine and Science

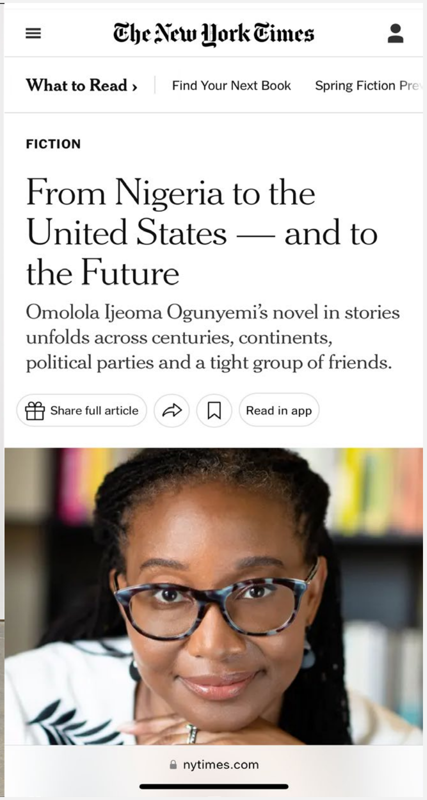
AIM-SWG

UCLA CDU CFAR
CENTER FOR AIDS RESEARCH

Introduction

- Born: University College Hospital, Ibadan, Nigeria

- Education:
  - B.A., Computer Science, Barnard College
  - M.S.E., Ph.D., Computer and Information Science, University of Pennsylvania



Charles R. Drew University of Medicine and Science

- CDU is a private, non-profit Historically Black Graduate Institution and a Hispanic Serving Institution
- Established 1966 in the aftermath of the Watts Rebellion in Los Angeles
- Mission: to graduate diverse health professional leaders dedicated to social justice and health equity for underserved populations
- CDU's Center for Biomedical Informatics (CBI)
  - established in 2007 to develop biomedical informatics solutions for medically underserved communities
  - funded by NIH grants, a CDU NIH endowment, state and foundation grants
  - CBI faculty members have backgrounds in computer science, internal medicine, sociology, and public health

Charles R. Drew University of Medicine and Science

- Located in Los Angeles' Service Planning Area 6 (SPA 6)
- SPA 6, one of eight SPAs, has
  - 1.1 million people of Latino (68%) and African American (28%) heritage
  - Disproportionately low number of hospitals, clinics and medical specialists
    - 32.5% of SPA 6 adults had difficulty accessing medical care vs. 26.3% county-wide*
    - HRSA-designated Health Professional Shortage Area
    - HRSA-designated Medically Underserved Area

*http://publichealth.lacounty.gov/ha/docs/2015lachs/keyindicator/ph-kih_2017-sec%20updated.pdf

Overview

- What is the healthcare safety net?
- Initial motivation for pursuing AI/ML work in the safety net
- Pros and cons of using EHR data for AI/ML in the safety net
- Diabetic retinopathy risk prediction study
- Current work on risk prediction for HIV

Safety Net

US Healthcare safety net

- includes
  - Federally Qualified Health Centers (FQHCs) and look-alikes
  - State and County hospitals
- provides healthcare
  - to ~34 million patients nationally
  - to ~6 million patients in California
  - regardless of patients' health insurance status or ability to pay

Safety Net

US healthcare safety net

- large number of patients compared to healthcare providers
- patients have often had gaps in care/limited healthcare access
- most safety net clinics/hospitals are unaffiliated with academic medical centers
- limited resources but must provide care to many
  - health information technology and informatics methods, including AI, can help
  - settings often don’t have staff with the HIT/informatics expertise

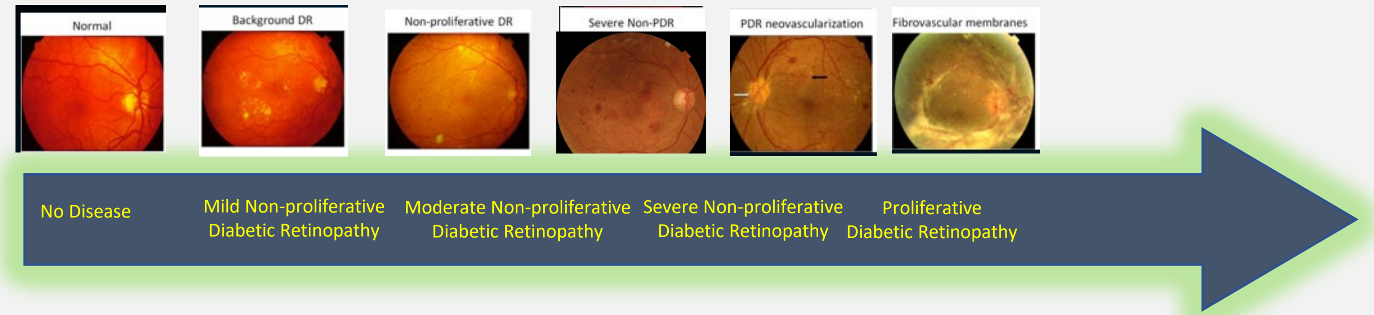
Teleretinal Screening and AI in Urban Safety Net Clinics

- Initial CDU safety net study in South Los Angeles
  - Challenge:
    - Insufficient numbers of ophthalmologists for timely annual in-person eye exams
    - Avoidable vision loss in diabetic patients
  - Proposed solutions:
    - 2010-2012: Assessed barriers to and facilitators of using **teleretinal screening** to detect diabetic retinopathy (DR) in 6 South Los Angeles safety net clinics (FQHCs)
    - 2016 – present: Machine learning (AI) on DR risk factors contained in EHR

Background

Diabetic retinopathy:

- Damage to blood vessels of the retina caused by excess blood glucose
- Left untreated can lead to blindness / treatable when detected early
- Leading cause of blindness in US adults aged 20 to 74 years
- A problem in medically underserved communities in the US



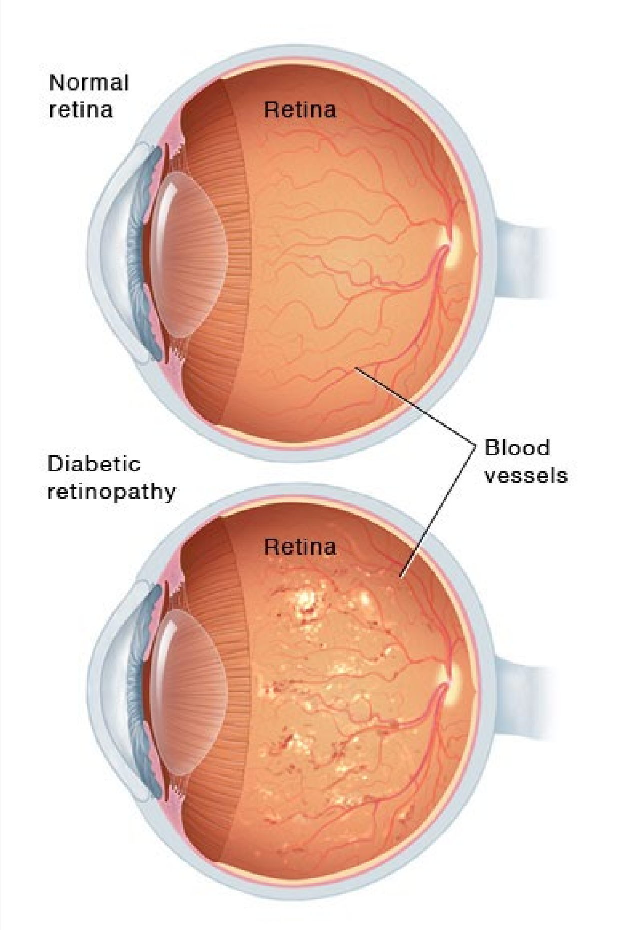
Images courtesy of Ricky Taira, PhD and Lauren Daskivich, MD

Background

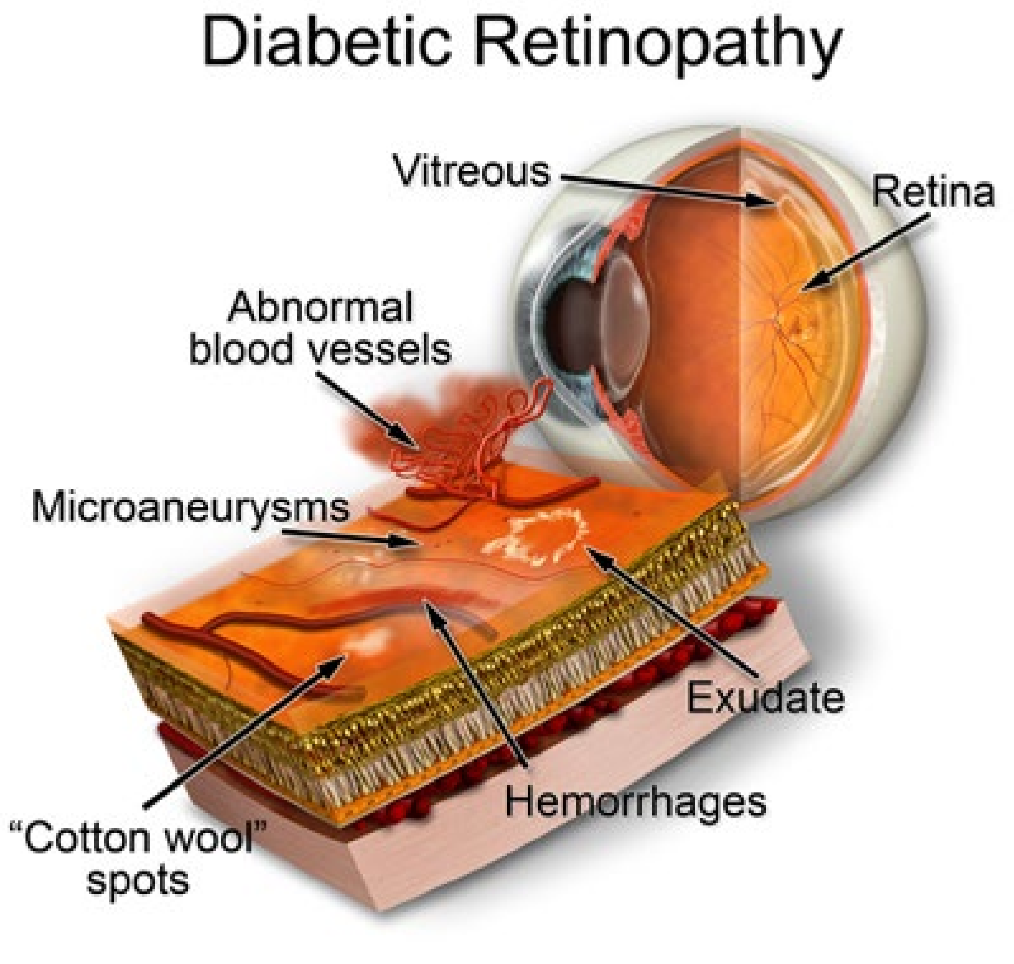
Diabetic Retinopathy (DR)

- Stages of DR
  - Mild non-proliferative diabetic retinopathy (NPDR)
  - Moderate NPDR
  - Severe NPDR
  - Proliferative diabetic retinopathy (PDR)
- Clinically significant macular edema (vision threatening and can occur at any stage of DR)

Background



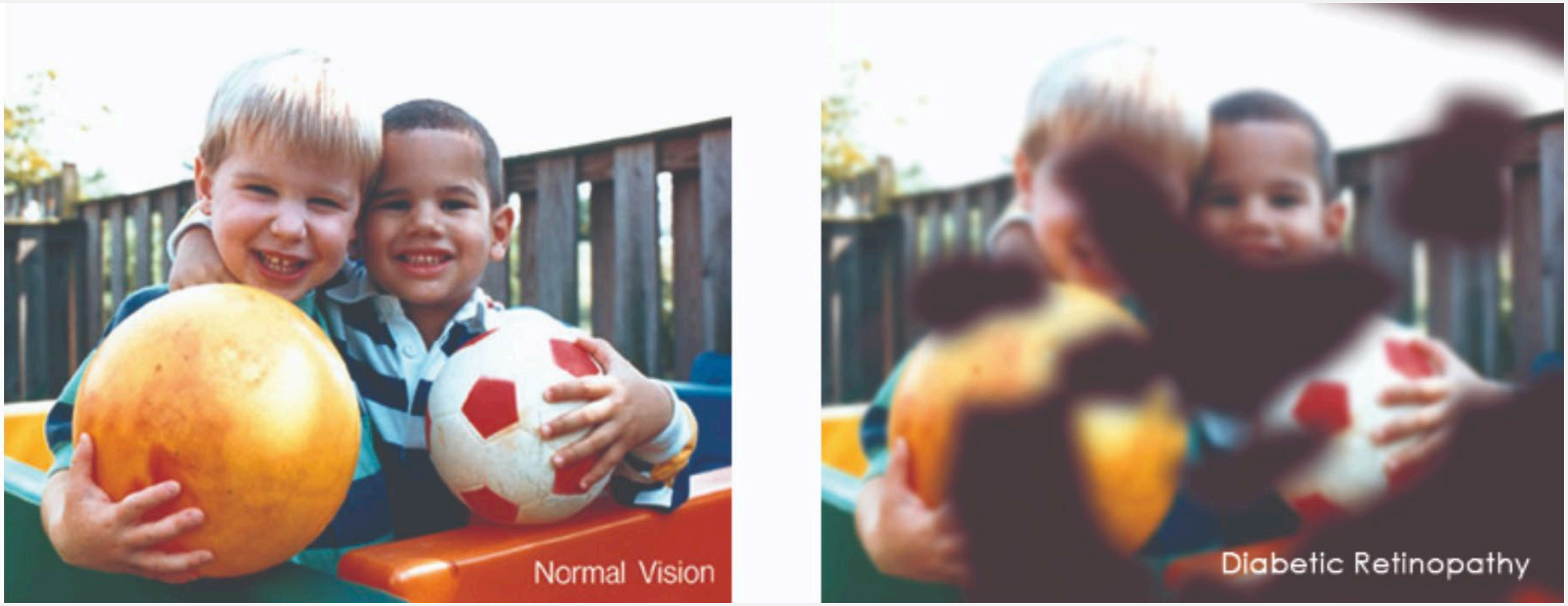
Normal retina
Retina
Diabetic retinopathy
Retina
Blood vessels



Diabetic Retinopathy
Vitreous
Retina
Abnormal blood vessels
Microaneurysms
Exudate
Hemorrhages
"Cotton wool" spots

Background

Partial vision loss from PDR bleed - total vision loss is irreversible



Images courtesy of the Discovery Eye Foundation

Background

Diabetic retinopathy risk factors from the biomedical literature:

- Duration of diabetes
- High blood glucose/poor blood sugar control
- Insulin treatment
- High blood pressure
- Dyslipidemia/high cholesterol
- Pregnancy
- Nephropathy
- Obesity
- Ethnicity
- Smoking

Observation: many of these risk factors are routinely collected in the EHR

Pros and Cons of using EHR data for AI/ML in the Safety Net

Pros:

- Safety net EHRs disproportionately represent
  - high risk and medically underserved populations
    - uninsured and underinsured patients
    - racial and ethnic minorities
    - immigrants, the unhoused, people with limited English proficiency
  - higher disease burden and chronic disease prevalence
    - may improve statistical power and model calibration
    - better for detecting rare but clinically important outcomes
  - real-world care complexity
    - irregular care patterns, social and structural barriers to care
    - capture real-world care patterns in AI/ML models versus idealized ones

Pros and Cons of using EHR data for AI/ML in the Safety Net

Pros (contd.):

- Safety net EHRs often collect information on
  - housing status
  - transportation needs
  - food insecurity
  - language preferences
  - other social determinants of health
- SDOH data can improve risk prediction and produce AI/ML models that support health equity
- bias reduction may result from training models on these data rather than on data from care settings with a lower disease burden
- AI/ML models developed on safety net data may have high public health and public policy relevance

Pros and Cons of using EHR data for AI/ML in the Safety Net

Cons:

- missing, incomplete, and inconsistent data
  - fragmented care across systems
  - incomplete medication histories
- increased bias for ML models if missingness is incorrectly modeled
- data reflect health access barriers in addition to biology
  - danger of predicting health system failure versus disease risk
- limited generalizability outside of safety net settings
- ethical and equity risks if data are used for rationing care rather than providing targeted support
- problem of making predictions without adequate resources for intervention

Artificial Intelligence

Machine Learning from EHR Data for Diabetic Retinopathy Detection

What is Artificial Intelligence?

- Computer systems or algorithms created to accomplish tasks that would ordinarily require human intelligence
- These tasks include
  - learning from data
  - recognizing patterns
  - making decisions
  - understanding natural language

- Personal view: AI in medical care should be to assist or augment healthcare professionals’ abilities, not replace them

What is Artificial Intelligence?

Examples of AI:

- machine learning (including deep learning)
- natural language processing (NLP)
- generative AI (large language models/LLMs)

Narrow (weak) AI

- designed or trained to perform specific tasks
- does not possess consciousness or general intelligence
- can perform better than a human expert in a narrow domain

What is Artificial Intelligence?

Examples of narrow AI in medicine:

- AI algorithms for medical imaging and diagnostics
- AI algorithms for predicting the risk of hospital readmission, sepsis, etc.
- Chatbots for patient appointment scheduling, symptom checking, medication management, mental health support, etc.
- PathAI – uses machine learning to help pathologists diagnose diseases from tissue samples
- Dragon Medical One – understands and transcribes clinicians' spoken language to update patient medical records
- Deep Genomics – uses AI to understand genetic mutations and develop targeted therapies for an individual
- Atomwise – uses machine learning to predict potential new drug candidates
- DaVinci surgical system robot – uses AI to assist surgeons with minimally invasive surgeries

What is Artificial Intelligence?

- Wide (general/strong) AI
  - describes a system that has generalized human cognitive abilities (artificial general intelligence)
  - can find a solution without human intervention
  - theoretical concept that is yet to be fully realized

Machine learning overview

Types of machine learning algorithms:

- Supervised learning
  - teach the computer how to learn by providing labeled examples with the "correct answers" that it can learn from
- Unsupervised learning
  - have the computer learn patterns by itself from unlabeled examples in a dataset
- Reinforcement learning
  - map from situations to actions by maximizing a reward/reinforcement signal (e.g., teaching a robot to avoid obstacles on a factory floor)

Machine learning overview

Supervised machine learning:

- Simple example - teach a computer how to recognize cats and dogs from lots of pictures
  - show the computer many pictures of cats and let it know they are cats
  - show the computer many pictures of dogs and let it know they are dogs
  - the computer tries to find patterns that make a cat different from a dog (e.g., cats have whiskers and dogs don't)
  - test the computer to see how well it learned by showing it a picture of a cat or dog it hasn't seen before and ask it to say whether the picture is of a cat or a dog
  - if the computer makes a mistake (calls a dog a cat or calls a cat a dog), we tell it the right answer, so it doesn't make that mistake again
  - as time passes, the computer gets better at recognizing cats and dogs because it keeps learning from the pictures and feedback it receives

- Medical example – breast cancer detection using convolutional neural networks from mammographic images labeled with benign or malignant tumors

Generative AI

- Prior to 2022, most AI used in the biomedical domain was non-generative AI that mostly utilized supervised learning
- Generative AI utilizes a combination of unsupervised learning and other techniques to create output – text, images, music, etc.
- Generative AI (large language models)
  - trained on huge amounts of unlabeled data
  - learn patterns and structure without being told what to look for
  - once trained, generate output by predicting what comes next in a sequence
  - may **hallucinate**
    - generate information that is plausible but incorrect or nonsensical
    - problematic in the healthcare domain where accuracy is crucial
    - human oversight is necessary in the healthcare domain – AI suggests and humans decide

AI benefits for patients and healthcare providers

- Enhanced diagnostic accuracy
- Personalized treatment plans for patients
- Automating routine administrative tasks to allow providers more time to focus on patients
- Improved patient monitoring for chronic conditions through wearable devices enhanced with AI
- Predictive analytics for crisis prevention and provision of proactive care

AI Concerns

- Data privacy and security
  - HIPAA-compliance is necessary in healthcare
- AI might make autonomous decisions
  - intended to augment/assist not replace humans
  - human oversight necessary in healthcare
- Lack of transparency
  - many successful AI algorithms are “black boxes”
- Fairness and bias
  - AI reflects the biases present in the data that it is trained on
  - AI developers may make claims about the generalizability of their tools that are not supported by reality
  - excitement over advances in AI in healthcare may overshadow the reality that benefits may not accrue to all patients

AI Concerns

Including social determinants of health in AI models may improve accuracy over utilizing clinical data alone

- Socioeconomic status
- Access to health services
- Access to healthy food
- Neighborhood and physical environment
- Educational attainment
- Health literacy
- Employment
- Experiences with discrimination
- Social support networks
- Community engagement

Original Investigation

July 6, 2022

Machine Learning–Based Models Incorporating Social Determinants of Health vs Traditional Models for Predicting In-Hospital Mortality in Patients With Heart Failure

Matthew W. Segar, MD, MS[1]; Jennifer L. Hall, PhD[2]; Pardeep S. Jhund, MBChB, MSc, PhD[3]; et al

Author Affiliations | Article Information

JAMA Cardiol. 2022;7(8):844-854. doi:10.1001/jamacardio.2022.1900

Key Points

Question Do machine learning (ML)–based models that incorporate social determinants of health (SDOH) improve the prediction of in-hospital mortality among patients with heart failure (HF)?

Findings In this cohort study, ML models developed in the Get With The Guidelines–Heart Failure (GWTG-HF) registry using race-specific and race-agnostic approaches were associated with an improvement in the prediction of in-hospital mortality after hospitalization for HF compared with the existing and rederived logistic regression models. The addition of SDOH was associated with an improvement in the performance and prognostic utility of the ML models in Black patients but not in non-Black patients.

Meaning The findings indicate that ML models incorporating SDOH may improve risk prediction of in-hospital mortality after hospitalization for HF, particularly in Black adults.

Segar MW, Hall JL, Jhund PS, Powell-Wiley TM, Morris AA, Kao D, Fonarow GC, Hernandez R, Ibrahim NE, Rutan C, Navar AM, Stevens LM, Pandey A. Machine Learning-Based Models Incorporating Social Determinants of Health vs Traditional Models for Predicting In-Hospital Mortality in Patients With Heart Failure. JAMA Cardiol. 2022 Aug 1;7(8):844-854. doi: 10.1001/jamacardio.2022.1900. PMID: 35793094; PMCID: PMC9260645.

AI Regulation

- European Union's AI Act passed by the EU Parliament on June 14, 2023, and formally adopted on March 13, 2024
  - "the first-ever comprehensive legal framework on AI worldwide"
  - focuses on privacy protections and data rights
  - establishes governing bodies
  - classifies AI risk as
    - unacceptable (e.g., cognitive behavioral manipulation, social scoring, compiling facial recognition databases via internet scraping/CCTV footage)
    - high
    - limited
    - minimal
  - **bans AI with unacceptable risk**

- US AI regulation is fragmented/sector specific – there is no overarching regulatory framework.
  - FDA regulates AI/ML used in medical devices ("software as a medical device")

ML for Diabetic Retinopathy Identification

ML for Diabetic Retinopathy Identification

Clinical Site: Los Angeles County Department of Health Services (LACDHS)

- Second largest municipal health care system in US
- Caters to ~750,000 unique patients a year
- ~142,000 patients are uninsured
- ~85,000 patients with diabetes seen between 2019 and 2020
- Teleretinal Diabetic Retinopathy Screening Program has ~10 optometrists with ophthalmologist overreads
- Teleretinal DR Screening Program objective is to screen all diabetic patients annually per DHS guidelines
- LACDHS DR screening rates improved from 37.7% in 2012 to 64% in 2019 after introduction of telehealth

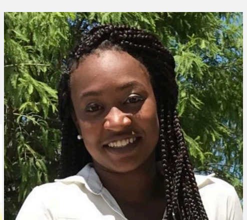
ML for Diabetic Retinopathy Identification

- Rationale:
  - ~36% of diabetic patients in LACDHS miss their teleretinal screenings/do not receive annual eye exams (~30,000 patients per year)
- Study Goal
  - Develop machine learning/ML methods to “detect” DR from data in electronic patient records
    - need to make DR risk assessment in the absence of digital retinal images
    - compare sensitivity, specificity and AUC of different ML approaches
    - identify and reach out to high-risk diabetic patients who skip teleretinal screening

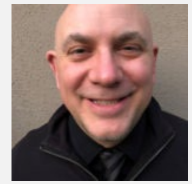
Team


Principal Investigators:
Omolola Ogunyemi, PhD (CDU)
Ricky Taira, PhD (UCLA)

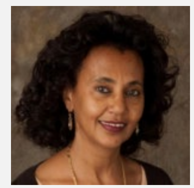
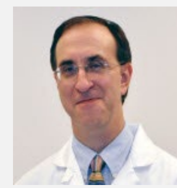
Sub-award principal investigator:
Lauren Daskivich, MD (LACDHS)



Project Coordinator:
Kyla Baron, MPH

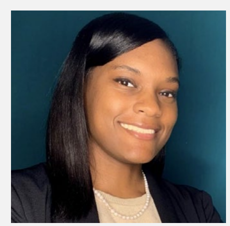




Co-Investigators:
Alex Bui, PhD (UCLA),
David Hindman, PhD (LACDPH),
Mayer Davidson, MD (CDU)

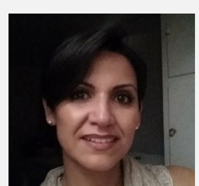
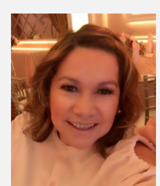




Co-Investigators:
Martin Lee, PhD (UCLA),
Senait Teklehaimanot, MPH, (CDU),
Robert Jenders, MD (CDU)

Software developer:
Meghal Gandhi, MS

Research Assistants:
Kevin Lopez, MS
Jasmine Jones

Health Educators:
Christina Martinez
Alma Coria

Evaluating machine learning classifiers

The usefulness/value of a binary classifier can be measured in several ways:

- Accuracy – percentage of all predictions that are correct (not always useful)
- Sensitivity/true positive rate (or recall) – measures the proportion of positive cases correctly identified as positive
- Specificity/true negative rate – measures the proportion of negative cases correctly identified as negative
- The trade-off between correctly identifying positives and correctly identifying negatives can be measured using the Area Under the ROC Curve (AUC)

Evaluating machine learning classifiers

–

Binary Classification Test (two by two table)		
	Test is negative	Test is positive
Patient does not have disease	TN	FP
Patient has disease	FN	TP

Sensitivity: proportion of correctly identified positives = TP/(TP+FN)
Specificity: proportion of correctly identified negatives = TN/(TN+FP)
Accuracy: proportion correctly identified = TP+TN/(TP+FN+FP+TN)
Area under the receiver operating characteristic (ROC) curve/

AUC: represents trade-off between sensitivity and specificity

Evaluating machine learning classifiers

Binary Classification Test (two by two table)		
	Test is negative	Test is positive
Patient does not have disease	850	50
Patient has disease	80	20

Accuracy: (850+20)/(850+50+80+20) = 87%
Sensitivity: 20/(20+80) = 20%
Specificity: 850/(850+50) = 94.4%

Evaluating machine learning classifiers

Binary Classification Test (two by two table)		
	Test is negative	Test is positive
Patient does not have disease	850	50
Patient has disease	80	20

This test has an accuracy of 87%, sensitivity of 20% and specificity of 94.4%.

Would you be comfortable using it as a test of whether people have this disease?

Why or why not?

Evaluating machine learning classifiers

Class imbalance in ML

- Outcome we want to predict is under-represented in the dataset we are learning from (minority class)
- This sometimes co-occurs with overlap in classes we are trying to separate
- Standard ML classifiers have a bias towards the class that has a greater number of instances in the data (majority class)
- Common problem with prediction in medicine
  - Health facility sees more patients who don’t have a particular disease/condition than patients who have it

Evaluating machine learning classifiers

Binary Classification Test (two by two table)		
	ML classifier says patient does not have disease	ML classifier says patient has disease
Patient does not have disease	900	0
Patient has disease	100	0

- Dataset with 900 negative training examples and 100 positive training examples
- We are trying to predict positive cases of disease
- Example of a class imbalance
- A ML classifier can obtain high <u>accuracy</u> by simply predicting that a patient does not have the disease.

Evaluating machine learning classifiers

Binary Classification Test (two by two table)		
	ML classifier says patient does not have disease	ML classifier says patient has disease
Patient does not have disease	900	0
Patient has disease	100	0

- Accuracy of ML classifier: 900/(900 + 0 + 100 + 0) = 90%
- Sensitivity: 0 / (100 + 0) = 0%
- Specificity: 900/(900 + 0) = 100%
- Classifier is useless at detecting this disease

Evaluating machine learning classifiers

Approaches to handling a class imbalance

- Modify learning algorithm to stress significance of correctly classifying the *minority* class
- Data pre-processing to rebalance the skewed distribution of data
  - Majority class undersampling
  - Minority class oversampling
- Cost-sensitive approaches that combine algorithm modifications and data pre-processing strategies
- Use of ensembles of classifiers that can increase the accuracy of classification by combining decisions/strengths of different individual classifiers

Methods

- Data source
  - LACDHS “ORCHID” Cerner EHR system
- Available data
  - Variables corresponding to known DR risk factors from biomedical literature
  - Variables suggested by clinician experts that address micro- and macro-vascular complications of diabetes
- Training and test set:
  - EHR records for Type I and Type II diabetic patients seen at LACDHS between 1/1/2015 and 12/31/ 2017
  - 40,631 total patients
    - 12,633 records of patients with DR (31.1%)
    - 27,998 records of patients with no DR (68.9%)
    - Dataset has a class imbalance

Methods

- Data source
  - LACDHS "ORCHID" Cerner EHR system
- External validation set:
  - EHR records for Type I and Type II diabetic patients seen at LACDHS between 1/1/2018 and 12/31/2018
  - No overlap between patients in training/test set and patients in external validation set
  - 9,300 total patients

Methods

Socio-demographic variables		
Age	Race	Ethnicity
Sex	Marital Status	Insurance Status
General Health Overview		
Diabetes Diagnosis Date*+	Date of Last Eye Examination	Pregnancy Status
Previous Diabetic Retinopathy Treatment	Smoking Status	Insulin Dependence
Clinical Measurements		
Body Mass Index	Diastolic Blood Pressure	Fasting Glucose Level
Blood Urea Nitrogen	Systolic Blood Pressure	HDL
Hemoglobin	Hemoglobin A1C	Triglycerides
Co-morbid Conditions		
Peripheral Vascular Disease	Hypertension	Stroke
Depression	Obesity	Nephropathy
Dyslipidemia	Neuropathy	Erectile Dysfunction
Condition of Interest		
Diabetic Retinopathy Diagnosis		

From: Ogunyemi OI, Gandhi M, Lee M, Teklehaimanot S, Daskivich LP, Hindman D, Lopez K, Taira R.
Detecting Diabetic Retinopathy through Machine Learning on Electronic Health Record Data from an Urban, Safety Net Healthcare System.
JAMIA Open. 2021 August 19;4(3):1 - 10.

Methods

Classification methods assessed:

- random forest (RF)
- support vector machine (SVM)
- extreme gradient boosting (XGBOOST)
- ensemble of classifiers
  - random forest, gradient boosting, and artificial neural networks
- deep neural network (DNN)

Data pre-processing methods applied to address class imbalance:

- majority class undersampling
- synthetic minority over-sampling technique (SMOTE)

Methods

Performed feature subset selection

Feature subset – 14 variables		
Age	Insulin Dependence	Ethnicity
Sex	Blood Urea Nitrogen	Hemoglobin A1C
Nephropathy	Diastolic Blood Pressure	Stroke
Neuropathy	Systolic Blood Pressure	Triglycerides
Hemoglobin	Duration of diabetes	

Methods

- Analyses performed using both R and Python
- Missing data handled with k-nearest neighbor imputation (k=9)
- Reserved a random selection of 34% of the dataset as a hold-out test set
- Performed 10-fold cross validation on remaining 66% of data
- Assessed best ML models from cross-validation process on test set and on external validation set

ML Results

Model Performance on 14 Variables with SMOTE on Test Set					
	RF SMOTE	Xgboost SMOTE	SVM SMOTE	Ensemble Model SMOTE	DNN SMOTE
Sensitivity	62.23%	49.89%	61.51%	63.93%	72.91%
Specificity	80.93%	86.55%	82.39%	79.84%	72.78%
AUC	0.8	0.781	0.797	**0.803**	0.8
Model Performance on 14 Variables with SMOTE on External Validation Set					
	RF SMOTE	XGBOOST SMOTE	SVM SMOTE	Ensemble Model SMOTE	DNN SMOTE
Sensitivity	57.24%	46.02%	58.13%	60.21%	70.63%
Specificity	83.43%	87.66%	84.36%	82.36%	74.54%
AUC	0.79	0.777	0.790	**0.795**	0.794

From: Ogunyemi OI, Gandhi M, Lee M, Teklehaimanot S, Daskivich LP, Hindman D, Lopez K, Taira R.
Detecting Diabetic Retinopathy through Machine Learning on Electronic Health Record Data from an Urban, Safety Net Healthcare System.
JAMIA Open. 2021 August 19;4(3):1 - 10.

ML Results

Model Performance on 14 Variables with Majority Class Undersampling on Test Set					
	RF Under	XGBOOST Under	SVM Under	Ensemble Model Under	DNN Under
Sensitivity	71.52%	70.85%	72.81%	70.68%	73.55%
Specificity	73.51%	74.61%	72.58%	74.96%	72.77%
AUC	0.799	0.800	0.798	0.803	**0.806**

Model Performance on 14 Variables with Majority Class Undersampling on External Validation Set					
	RF Under	XGBOOST Under	SVM Under	Ensemble Model Under	DNN Under
Sensitivity	69.06%	66.78%	70.00%	67.38%	72.17%
Specificity	76.01%	77.35%	75.24%	77.09%	74.20%
AUC	0.791	0.792	0.794	0.794	**0.8**

From: Ogunyemi OI, Gandhi M, Lee M, Teklehaimanot S, Daskivich LP, Hindman D, Lopez K, Taira R. Detecting Diabetic Retinopathy through Machine Learning on Electronic Health Record Data from an Urban, Safety Net Healthcare System. JAMIA Open. 2021 August 19;4(3):1 - 10.

Results

DRRisk: The Diabetic Retinopathy Risk Assessment Tool

DRRisk is an educational tool that assesses the risk of current diabetic retinopathy in individuals who have diabetes. It uses fourteen risk factors to make a determination of an individual's current risk of retinopathy.

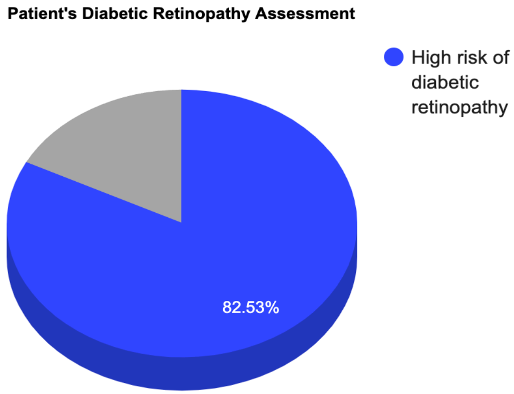
1. Insulin dependence
2. BUN
3. Systolic blood pressure
4. Neuropathy
5. Hemoglobin A1c
6. Hemoglobin
7. Sex
8. Ethnicity
9. Nephropathy
10. Duration of diabetes
11. Triglycerides
12. Stroke
13. Diastolic blood pressure
14. Age

Assess Diabetic Retinopathy

The diabetic retinopathy risk assessment tool was developed at the Center for Biomedical Informatics at Charles R. Drew University of Medicine and Science, using electronic health record data from 27, 223 Los Angeles County Department of Health Services (LACDHS) patients with Type 1 or Type 2 diabetes seen between 2015 and 2017. The average age of the individuals was 58 years, 57.5% of the individuals were women and 75.3% of the individuals were Latino. The tool is based on a deep neural network learned on the LACDHS data. Details can be found in the following publication:

Ogunyemi OI, Gandhi M, Lee M, Teklehaimanot S, Daskivich LP, Hindman D, Lopez K, Taira R. Detecting Diabetic Retinopathy through Machine Learning on Electronic Health Record Data from an Urban, Safety Net Healthcare System. JAMIA Open. 2021 August 19;4(3):1 - 10.[Click here to view the publication]

DRRisk: The Diabetic Retinopathy Risk Assessment Tool



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	55 years
2. Duration of Diabetes(Years)	5 year(s)
3. Systolic Blood Pressure	None mmHg
4. Diastolic Blood Pressure	None mmHg
5. Blood Urea Nitrogen(BUN)	25.0 mg/dL
6. Hemoglobin	None g/dL
7. Hemoglobin A1C	8.0 %
8. Triglycerides	None mg/dL
9. Sex	M
10. Ethnicity	Hispanic or Latino
11. Insulin Dependence	N
12. Neuropathy	None
13. Nephropathy	None
14. Stroke	None

Work in Progress

- Examining effectiveness of AI/ML-driven patient outreach to LACDHS pandemic strategy of HbA1C > 9
  - 10,731 of 31,072 patients who had missed eye exam were seen at clinic within a one-year period
  - Initial results on those patients show AI/ML better at detecting vision threatening retinopathy:
    - Severe NPDR
    - PDR and
    - Clinically Significant Macular Edema at any stage of DR
  - Manuscript in progress
- Prelude to clinical trial & implementation science study examining AI/ML driven outreach versus status quo

Acknowledgments

This work was funded by the National Library of Medicine under grant 1 R01 LM012309.

The content presented is solely the responsibility of the investigators on that grant and does not necessarily represent the official views of the National Institutes of Health.

Work in Progress

Ongoing study

Background: In Los Angeles County (LAC), the virologic suppression goal is 95% suppression, but the actual suppression rate is 64%. Identifying individuals at risk for virologic failure gives clinicians an opportunity to modify existing treatment approaches, and AI/ML provides a potential tool to achieve this.

Goal 1: Identify patients in South Los Angeles who are initiating anti-retroviral therapy (ART) for the first time, who are at risk for virologic failure

Background: The Los Angeles County 2023 annual surveillance report notes a retention rate of 51% for those diagnosed with HIV and was lowest among women, those aged 20-49 years, Blacks, and injection drug user transmission categories.

Goal 2: Identify patients in South Los Angeles at risk of dropping out of HIV care by integrating social determinants of health (SDOH) data, clinical records, and behavioral metrics

Study PIs: Lola Ogunyemi, LaShonda Spencer
Funding: American Academy of HIV Medicine Caceres Award

Thank you!

Questions?

Contact info: lolaogunyemi@cdrewu.edu