



Artificial Intelligence in Healthcare

Mose Wintner, Ph.D.
Data Scientist
Kaiser Permanente Health Innovation Studio
moses.a.wintner@kp.org
September 15, 2018
35th Annual SCPMG Ophthalmologist Symposium



Outline

- What are AI, ML, deep learning?
 - Supervised vs. unsupervised machine learning
- Neural networks crash course
- General survey of recent AI developments in healthcare
 - Genomics & personalized medicine
 - AI-assisted surgery
 - AI-powered nursing assistants
 - Administrative workflow automation
 - Image analysis for prognostics risk assessment and diagnostic support
- Interpreting black-box predictions
- Questions

What are AI, ML, deep learning?



What are AI, ML, deep learning?



- **Artificial intelligence (AI)** involves machines that can perform tasks that are characteristic of human intelligence." – John McCarthy

What are AI, ML, deep learning?




- **Artificial intelligence (AI)** involves machines that can perform tasks that are characteristic of human intelligence." – John McCarthy
- **Machine learning (ML)** is a common mathematical ingredient of AI.


What are AI, ML, deep learning?




- **Artificial intelligence (AI)** involves machines that can perform tasks that are characteristic of human intelligence." – John McCarthy
- **Machine learning (ML)** is a common mathematical ingredient of AI.
- **Deep learning** usually refers to a complex family of ML algorithms called *artificial neural networks*, whose architecture is inspired by that of the brain.

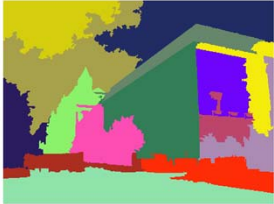
Unsupervised vs. supervised ML 

UNSUPERVISED	SUPERVISED
No human feedback or instructions necessary for correct performance	Human feedback / instructions necessary for correct performance, e.g. anything requiring expert input, training data, etc.

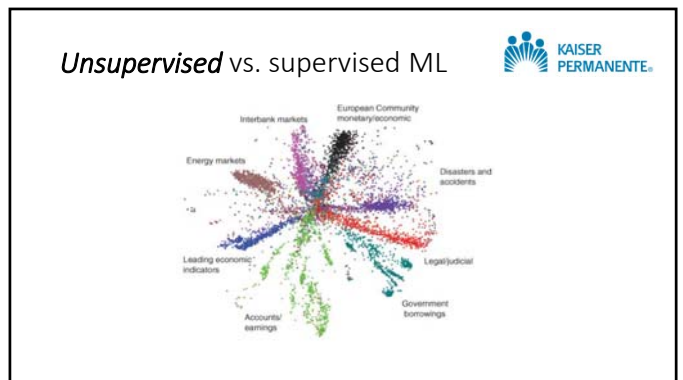
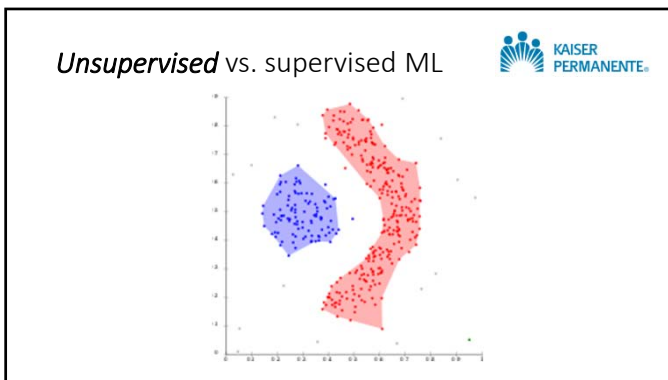
Unsupervised vs. supervised ML 




(a)

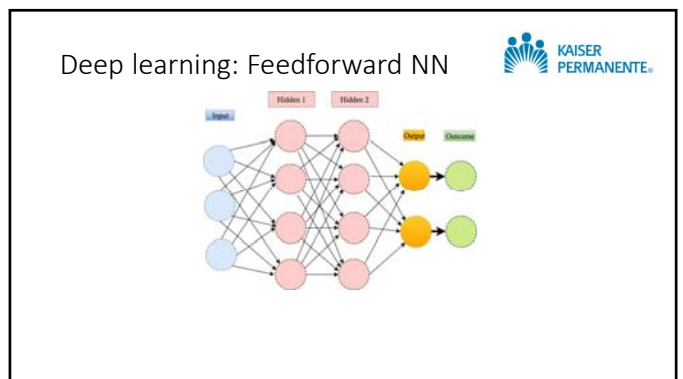


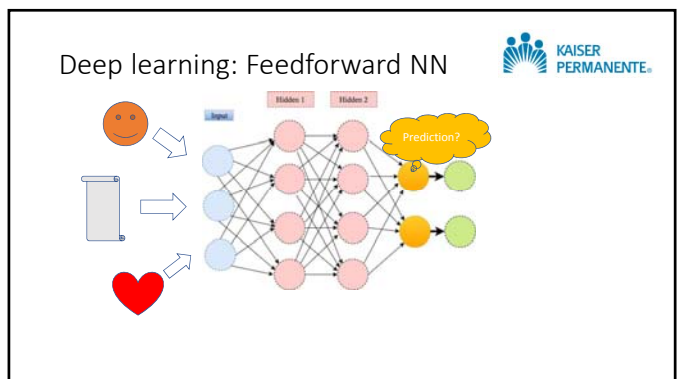
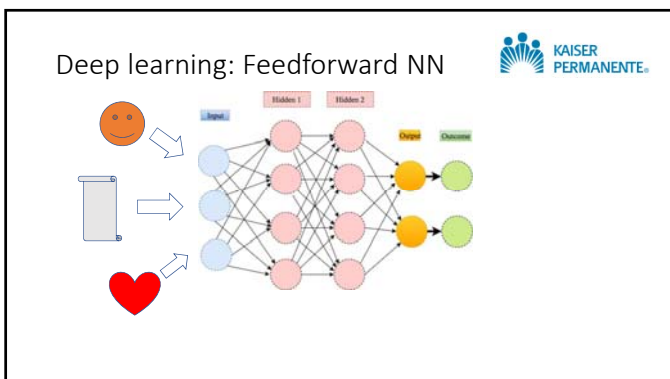
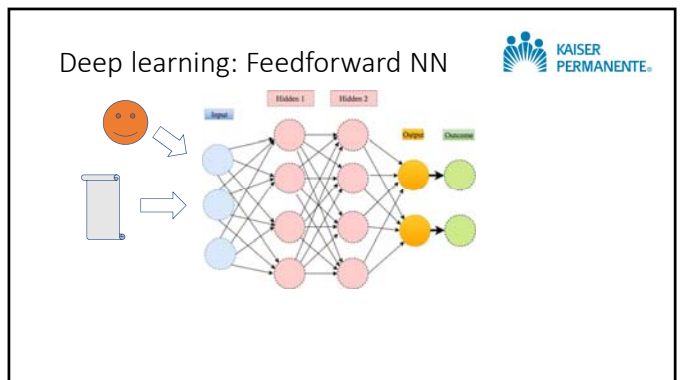
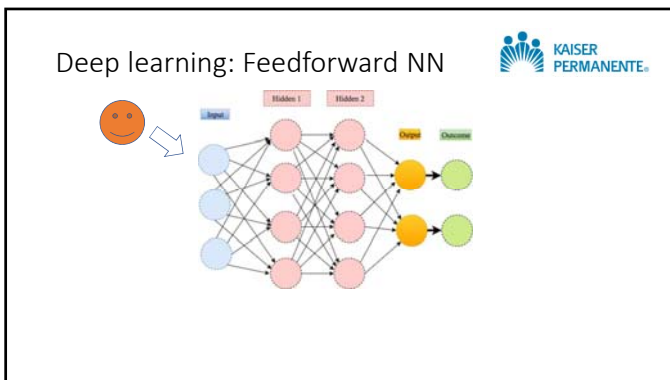
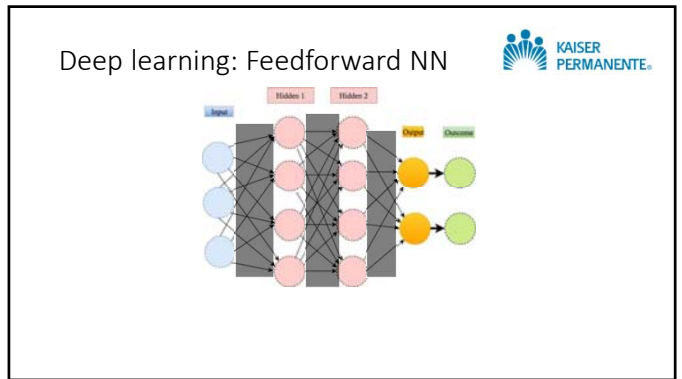
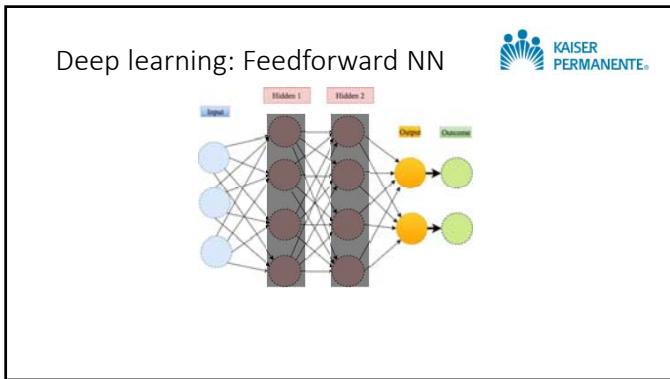
(b)

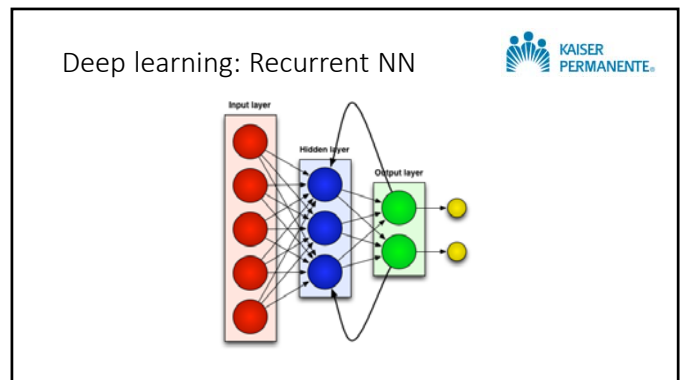
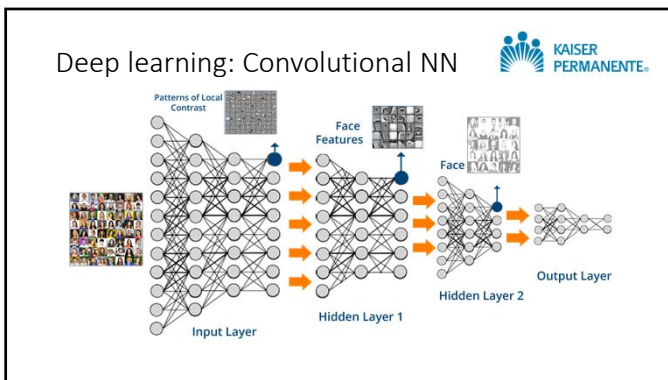
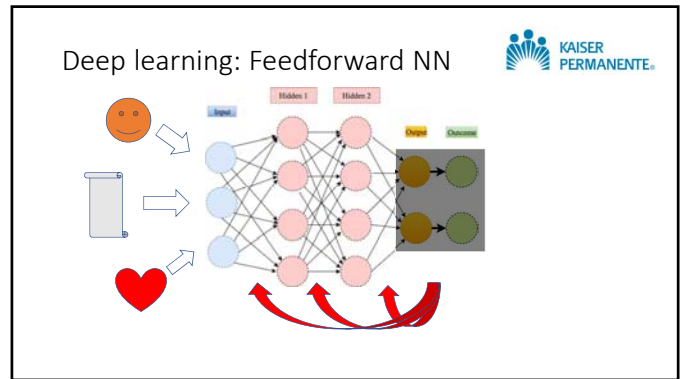
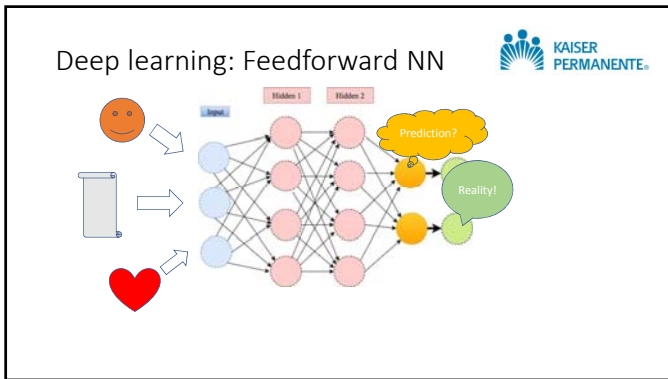


You won't be out of a job 

- Almost every application of ML to medicine is supervised
- Most medical tasks are still too complex for AI
- Patients want to be treated by a human, not a computer
- Doctors needed for medical AI
 - Instruct developers to include necessary indicators
 - Provide clinical insight
 - Maintain integrity of algorithm
 - Provide care

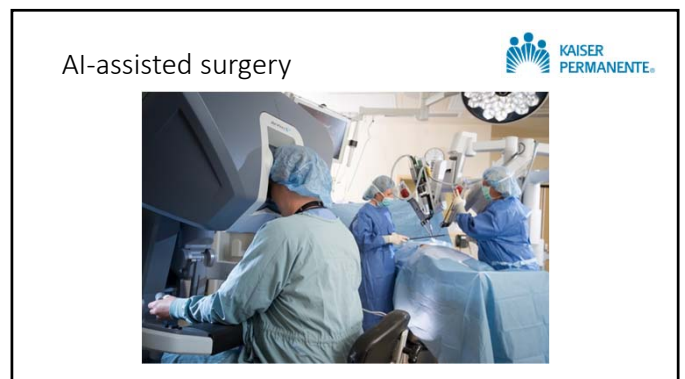





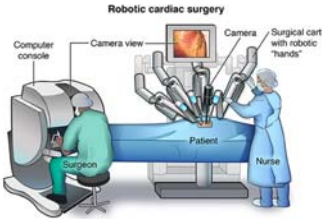


Genomics & PM: 2016-Present


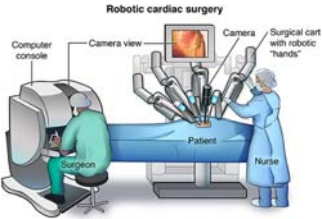
- (2016) Google's DeepVariant automatically identifies small mutations in sequencing data
- (2016) "Smart dosage" using phenotypic data for liver transplant immunosuppression
- (2017) Over 12m consumers worldwide take genetic health marker and/or ancestry tests, more than all previous years combined



AI-assisted surgery


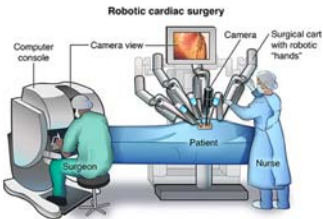



AI-assisted surgery


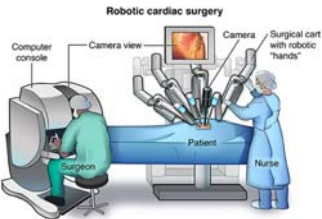
80% reduction
in surgical complications

AI-assisted surgery






Reductions could result in **21% shorter** post-surgery hospital stay on average

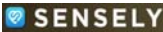

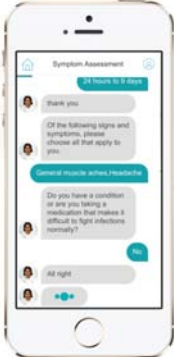
AI-assisted surgery

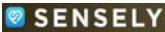
Up to **\$40B in annual savings** nationwide by 2026







AI-powered nursing assistants

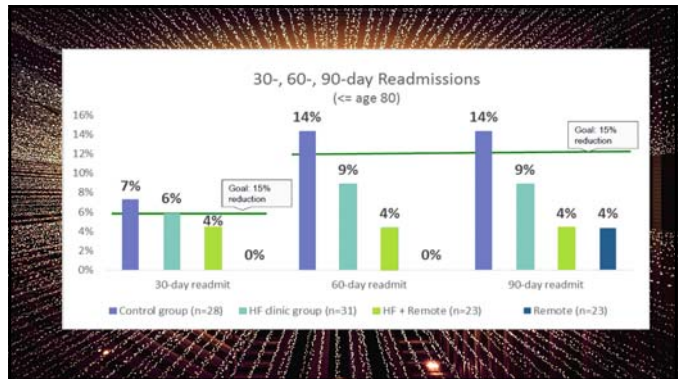




AI-powered nursing assistants








AI-powered nursing assistants





Administrative tasks and workflow





AI for Coordinated Care




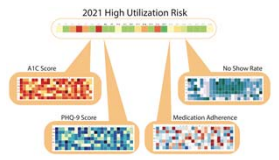
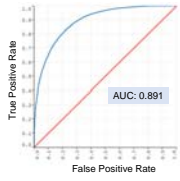
A digital care coordination tool that assists care managers by:

- 1 Generating a personalized care plan based on a patient's needs and goals
- 2 Identifying patients at risk of deviating from their care plan and prioritizing outreach
- 3 Identifying patterns and recommending the best course of action and education for each individual patient

Goals of the experiment

- Maximize**
 - Care Pathway adherence
 - Care manager efficacy
- Minimize**
 - Provider burden
 - Disease progression
 - Complications and acute care utilization
 - Missed appointments, med, and screenings
 - Duplicative testing



High utilizer prediction

Current Pilots

- West LA – Using prediction to target social needs assessment
- Pan City – Using the prediction to optimize social services referrals

CHF Predictions





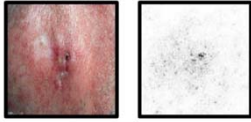
Readmission Impact	End of Life Prediction Impact	Reduced EF Prediction Impact
168 extra Readmissions over LACE	228 Patients predicted mortality within 6M-3Y	378 Patients detected at HF-EF 5 years ahead
962 Inpatient Days	Admissions: 388 Readmissions: 101 ED Visits: 567 Outpatients visits: 11,557	Admissions: 999 Readmissions: 140 ED Visits: 2000 Outpatients visits: 79,126

Count of Correctly Predicted 30d Readmissions: **1781**

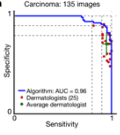
Cost of Correctly Predicted 30d Readmissions: **>20M**

Skin cancer detection from smartphone photos (2017)

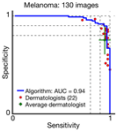




a


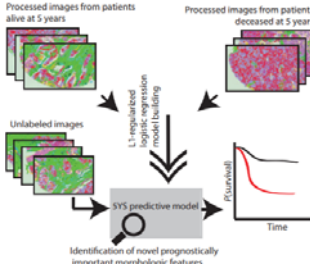


Algorithm: AUC = 0.96
Average dermatologist




Algorithm: AUC = 0.94
Average dermatologist

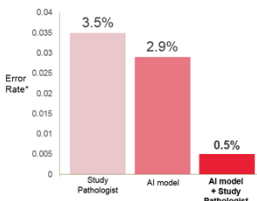
ID of metastasized breast cancer (2016)

ID of metastasized breast cancer (2016)




(AI + Pathologist) > Pathologist



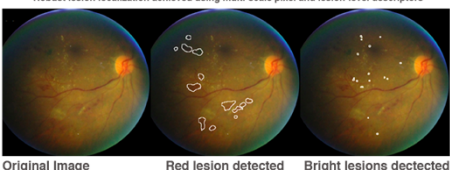
*Error rate defined as 1 - Area under the Receiver Operator Curve
**A study pathologist, blinded to the ground truth diagnoses, independently scored all evaluation slides.
© 2016 PathAI

Diabetic retinopathy early detection (2016)


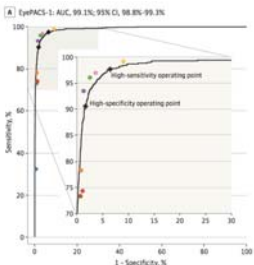


Lesion localization


Robust lesion localization achieved using multi-scale pixel and lesion-level descriptors



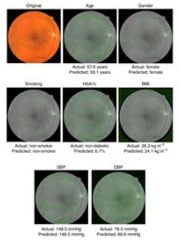
Diabetic retinopathy early detection (2016)


Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning (2018)



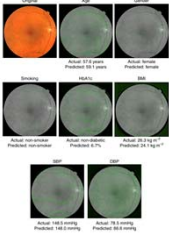
- Predicted cardiovascular risk factors not previously detected in retinal images, including




Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning (2018)



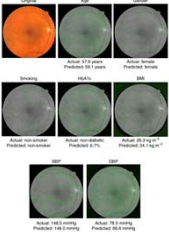
- Predicted cardiovascular risk factors not previously detected in retinal images, including
 - Age (within 3.26 years on average)




Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning (2018)



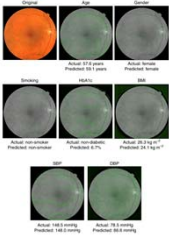
- Predicted cardiovascular risk factors not previously detected in retinal images, including
 - Age (within 3.26 years on average)
 - Gender (0.97 AUROC)




Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning (2018)



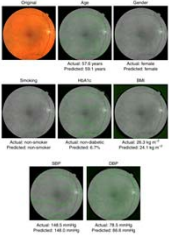
- Predicted cardiovascular risk factors not previously detected in retinal images, including
 - Age (within 3.26 years on average)
 - Gender (0.97 AUROC)
 - Smoking status (0.71 AUROC)




Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning (2018)



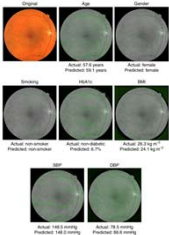
- Predicted cardiovascular risk factors not previously detected in retinal images, including
 - Age (within 3.26 years on average)
 - Gender (0.97 AUROC)
 - Smoking status (0.71 AUROC)
 - HbA1c (within 1.39% points on average)




Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning (2018)



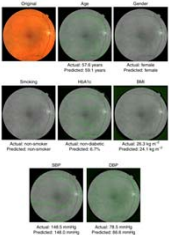
- Predicted cardiovascular risk factors not previously detected in retinal images, including
 - Age (within 3.26 years on average)
 - Gender (0.97 AUROC)
 - Smoking status (0.71 AUROC)
 - HbA1c (within 1.39% points on average)
 - Systolic blood pressure (within 11.23 mmHg on average)

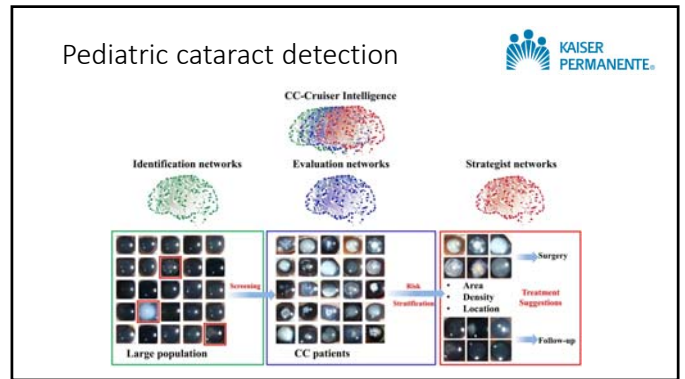
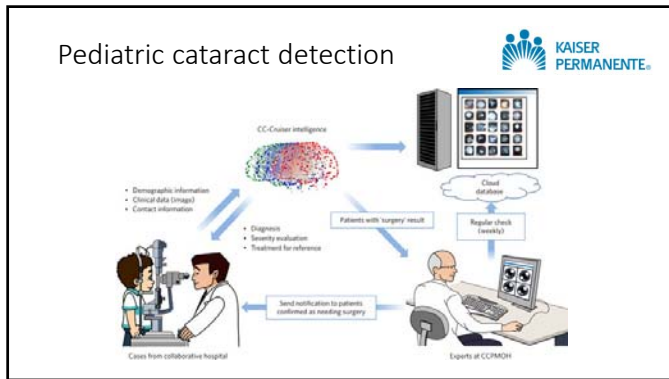


Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning (2018)



- Predicted cardiovascular risk factors not previously detected in retinal images, including
 - Age (within 3.26 years on average)
 - Gender (0.97 AUROC)
 - Smoking status (0.71 AUROC)
 - HbA1c (within 1.39% points on average)
 - Systolic blood pressure (within 11.23 mmHg on average)
 - History of major adverse cardiac events (0.70 AUROC)





AMD classification

KAISER PERMANENTE

- Age-related Macular Degeneration classification (dry/wet)
 - 2.6m OCT images from 43k scans, 9k patients
 - 93% accuracy
 - High sensitivity: 92.6%
 - High specificity: 93.7%

AMD treatment referral

KAISER PERMANENTE

- Evaluating AMD treatment need with OCT
 - 183k OCT B-scans
 - 95.5% accuracy
 - High sensitivity: 90.1%
 - High specificity: 96.2%

Deep learning with EHR (2018)


KAISER PERMANENTE

- Makes predictions from patient EHR upon hospital admission



Deep learning with EHR (2018)


KAISER PERMANENTE

- Makes predictions from patient EHR upon hospital admission
- All-cause 30 day readmission prediction: AUROC = 0.87



Deep learning with EHR (2018) 


- Makes predictions from patient EHR upon hospital admission
- All-cause 30 day readmission prediction: AUROC = 0.87
- Inpatient mortality prediction: AUROC = 0.97

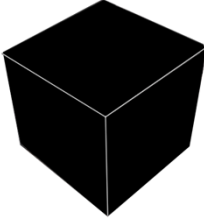




Deep learning with EHR (2018) 

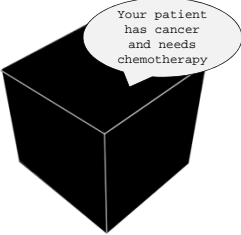
- Makes predictions from patient EHR upon hospital admission
- All-cause 30 day readmission prediction: AUROC = 0.87
- Inpatient mortality prediction: AUROC = 0.97
- Primary Diagnosis prediction: Recall@5 = 0.88





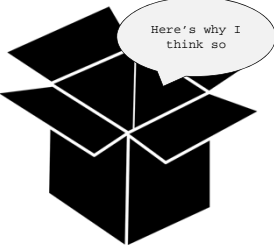
Deep learning: Interpretability 




Deep learning: Interpretability 

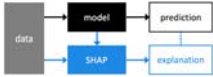


Deep learning: Interpretability 



SHAP (SHapley Additive exPlanation) 

SHAP claims "A unified approach to explain the output of any machine learning model."

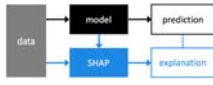


```

    graph LR
      data --> model
      model --> prediction
      prediction --> explanation
      SHAP --> explanation
  
```

SHAP (SHapley Additive exPlanation)

SHAP claims "A unified approach to explain the output of any machine learning model."

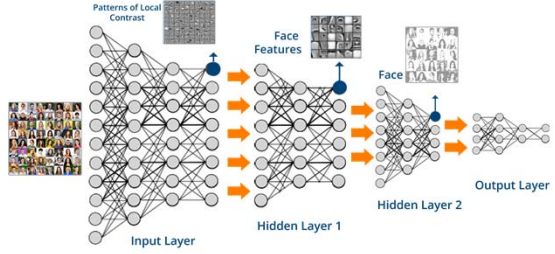


base value: 24.41

higher > lower

PTTRATIO = 15.3 LSTAT = 4.98 RM = 6.575 NOX = 0.538 AGE = 65.2 RAD = 1

Deep learning: Convolutional NN



Input Layer Hidden Layer 1 Hidden Layer 2 Output Layer

DeepMind Segmentation-Classification

Traditional AI

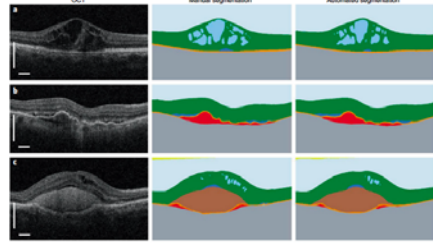
INPUT (image) → BLACK BOX → OUTPUT (Diagnosis)

DeepMind's Framework

INPUT (image) → Segmentation network → TISSUE MAP (with labels for anatomy, pathology, artefacts) → Classification network → OUTPUT (Referral suggestion + Diagnosis)

Referral suggestion (%)	Diagnosis probability (%)
Observation only	Normal 77.1
Observation only	CMV 0.0
Observation only	MIO 1.4
Observation only	Full mac. hole 11.0
Observation only	Full mac. hole 24.2
Observation only	CGM 15.0
Observation only	MET 66.4
Observation only	ERM 0.000
Observation only	Sec. atrophy 51.9

DeepMind Segmentation-Classification



OCT Manual segmentation Automated segmentation

- vitreal or subhyaloid cysts
- Posterior hyaloid
- Epiretinal membranes
- Neovascularized retina
- intra-retinal fluid
- Subretinal fluid
- Subretinal hyper reflect. mat.
- Retinal pigment epithelium
- Choroidal PED
- Serous PED
- Choroidal and outer layers
- Flaking artefact
- Slit-like artefact
- Roll-over artefact

Conclusions

- Great strides are being made in image analysis, especially retinal fundus photography
- AI for diagnostics are still in their infancy
- AI is a *tool* for medical care, not a replacement for it

Diagnostics & clinical judgment

Smarter tools for making unbiased, evidence-based clinical judgments on the part of physicians and patients alike

Use of AI in diagnostics is still in its infancy, but many promising results

- Statistical models to infer patient-specific probabilities of disease outcomes, including mortality, complications, care utilization
- NLP (natural language processing) for extraction of information from clinical notes and medical literature
- Tools for direct diagnosis

AI @ KP



- Collecting patient data electronically for over a decade
- 44 petabytes of data = 170,000 physician-years of experience
- Capability for massive retrospective studies and hypothesis generation



1.4 Million iPhones



30 stacks of CD-ROMS as tall as Mt. Whitney



Enough floppy disks to circle the globe 65 times!

SHAP (SHapley Additive exPlanation)



- Assigns each variable/feature a **feature attribution** value for each individual model prediction $f(x)$, where

$$f(x) \approx g(z') = \phi_0 + \sum_{i=1}^p \phi_i z'_i$$

- x is the model input data
- p is the number of variables/features in the full model
- z'_i is a binary variable encoding whether variable i is used in the model
- z' is a vector containing the z'_i
- $g(z')$ is a parallel "explanation model" which agrees locally with the prediction
- ϕ_i is the feature attribution for variable i in the model's prediction for data x

Genomics & Personalized Medicine



The branch of molecular biology concerned with the structure, function, evolution, and mapping of genomes.

"The gap that is currently blocking medicine right now is in our inability to accurately map genetic variants to disease mechanisms and to use that knowledge to rapidly identify life-saving therapies."

– Brendan Frey, CEO of Deep Genomics
MIT Technology Review, December 4, 2017

Genomics & PM: Selected history



- (2003) Human Genome Project officially completed
- (2005) Map of human genetic variation (HapMap) published in *Nature*
- (2006) 23andme, the first direct-to-consumer DNA ancestry testing firm, founded
- (2007) Explosion of 150+ published replicated disease-associated genetic sequence variants, up from single-digit # published in 2006
- (2009) First comprehensive analysis of cancer genomes published
- (2013) US Supreme Court rules that human DNA cannot be patented because it is a "product of nature"

Prognostics: History



- (1972) Cox proportional hazards model
- (late 1980s) Nascency of AI systems for medical prognostics/diagnostics
- (early 1990s) Rule-based expert AI systems and Bayesian networks appear in medicine
- (1990s) Further applications of decision trees and engineering logic in medical prognostics and diagnostics
- (early 2000s) First uses of artificial neural networks (i.e. deep learning) in prognostics