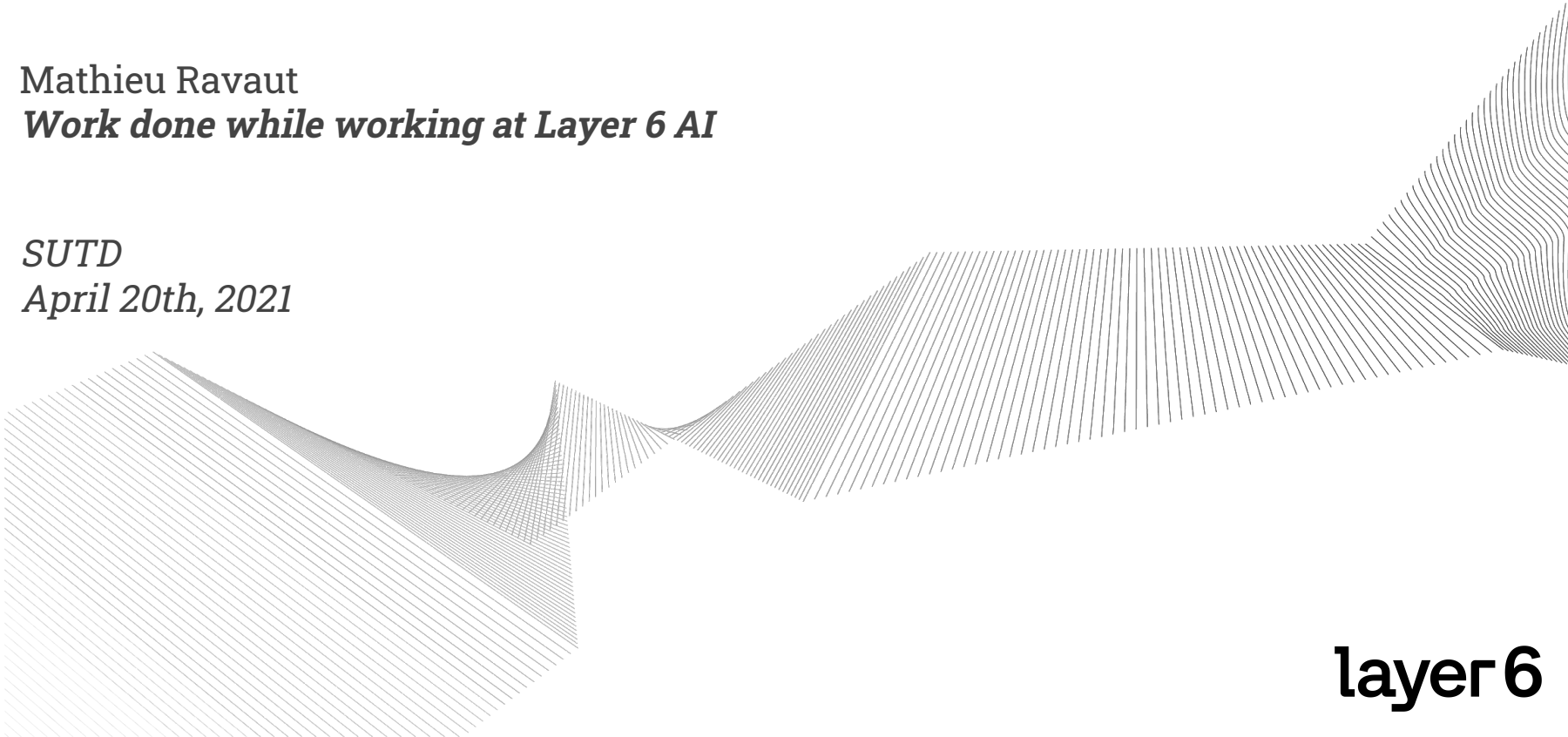
Tackling diabetes with machine learning

Mathieu Ravaut
Work done while working at Layer 6 AI

SUTD
April 20th, 2021



layer 6

About me

- [FRANCE] M.Eng, Applied Mathematics @ Ecole Centrale Paris (2014 - 2018)
- [SG] ML Research Intern @ Thales Solutions Asia, @ A-STAR (2016 - 2017)
- [CANADA] MSc Applied Computing @ University of Toronto (2017 - 2018)
- [CANADA] ML Research Scientist at Layer 6 AI, Toronto (2018 - 2020)
  - The AI lab of TD Bank, Canada's second largest bank
  - Projects on ML & NLP for healthcare, insurance, recommendation systems
- [back in SG!] NLP PhD candidate @ NTU (2021 -) w Prof. Shafiq Joty
- … love machine (& deep) learning <3

layer 6

Goal of my talk

Give an idea of how an **applied research** project in machine learning is done **end-to-end** in industry.
Machine (& deep) learning is talked about everywhere these days, but what is it really all about?
What can machine (& deep) learning do, how does it bring value?

Goal of my talk

Give an idea of how an **applied research** project in machine learning is done **end-to-end** in industry.
Machine (& deep) learning is talked about everywhere these days, but what is it really all about?
What can machine (& deep) learning do, how does it bring value?

In this case, we tackle the following question:
How can machine (& deep) learning be leveraged to make a change against diabetes with the data collected by the public healthcare system (specifically, in Ontario, Canada)?

Goal of my talk

Give an idea of how an **applied research** project in machine learning is done **end-to-end** in industry.
Machine (& deep) learning is talked about everywhere these days, but what is it really all about?
What can machine (& deep) learning do, how does it bring value?

Challenges:

- Patient-level health data is **anonymized** and **access is (very) restricted.**
- Data is **not clean**: missing values, non-consistent codes, etc
- Data is **not prepared** for machine (& specifically deep) learning.

Goal of my talk

Outcomes:

- *Type-2 Diabetes Onset Prediction at the Population-Level using Machine Learning and Routinely Collected Administrative Health Data*, Ravaut et al. (**Jama** Network Open, 2021)

- *Predicting Adverse Outcomes due to Diabetes Complications with Machine Learning Using Administrative Health Data*, Ravaut et al. (**Nature** NPJ Digital Medicine, 2021)

layer 6

Outline

- A. Motivation.
- B. Task.
- C. Data.
- D. Framework.
- E. Model.
- F. Results.
- G. Example.

layer 6

A. Motivation

- Diabetes is the one of the **most prevalent** chronic diseases in the world with **~8.8%** prevalence rate (10% in Ontario, 10.5% in Singapore)

- **Dramatic complications** severely impairing one's life:
  - Amputation
  - Blindness
  - Kidney failure
  - ...

- That makes diabetes the **costliest** chronic disease for Ontario.

- If spotted early enough, these complications can be **avoided**.
  - Better patient outcome
  - **Save costs** for the Ontario province.

layer 6

A. Motivation

Quick estimation of the total yearly cost of diabetes complications hospitalizations:

Complication	**Cases per year in Ontario (approx)**	**Average cost per hospitalization ($CAD))**	**Total cost (M $CAD))**
Hyper/hypo - glycemia	6,790	10,534	71
Tissue infection	36,063	21,868	789
Retinopathy	2,563	13,333	34
Cardiovascular	64,354	17,863	1,150
Amputation	19,976	9,568	191

Total: ~2.24B $CAD /year !

A. Motivation

- **Machine learning** is said to enable **personalized** medicine.
- But **where/how** can machine learning be applied to diabetes?

A. Motivation

- **Machine learning** is said to enable **personalized** medicine.
- But **where/how** can machine learning be applied to diabetes?
- Use cases
  - Diabetes onset prediction
    Predict whether the patient is going to get diabetes (binary class.)
  - Diabetes complications prediction
    Predict whether the patient is going to get a set of complications (several binary class.)

A. Motivation

- There are already existing data science studies on diabetes.
- What is specific here?
    - Go beyond models focusing on a small subgroup and **include the full population.**
      Very little exclusions.
    - **Very large scale:** millions of patients, across more than 10 years.
      A recent study published in Nature had 900 patients.
    - **Deploy** the diabetes onset and the diabetes complications systems in **Ontario.**

layer 6

B. Task

- Diabetes onset prediction: **binary** classification

- Diabetes complications: **multi-label** binary classification
  We identify **severe outcomes** of diabetic complications:
  - Hospitalizations
  - Ambulatory usage

layer 6

B. Task

But what are the **labels**?

- Unlike on Kaggle, **labels were not given**.
   We had to build them ourselves.

- We flag **ICD-10** codes:
  - Representing diabetes: find the earliest one
  - Corresponding to hospitalizations due to a diabetes complication

- Inherent noise:
  - Choice of codes
  - Reporting error in the database
  - Billing bias
  - A code is not the complication itself

B. Task

- We predict the **incidence** of diabetes/complications at each **quarter.**

- We pick **5 complications:**
  - Severe hyper or hypoglycemia
  - Tissue infection
  - Retinopathy
  - Cardiovascular event (stroke, heart failure)
  - Amputation

➡ Cover all the main complications

B. Task - statistics

Disease	Quarterly incidence (%)
Diabetes	0.21
1 - Hyper/hypo - glycemia	0.14
2 - Tissue infection	0.65
3 - Retinopathy	**0.04**
4 - Cardiovascular	1.09
5 - Amputation	0.35

Note the **extreme sparsity** of some complications...

layer 6

C. Data - cohort

- Need to use **2 different cohorts** (one for each task)

- To predict **diabetes**:
  - Purely random sample of 3m people
  - Negative examples are patients not developing diabetes in the future

- To predict **diabetes complications**
  - Everyone developing diabetes at some point: ~2m
    Goal is to include **as many people** as possible.

- **Exclusions** if one of the following is broken:
  - Alive as of January 1st, 2012
  - Date of Last Contact (DOLC) after the target window
  - Landing Date in Canada before the end of the observation

layer 6

C. Data - source

- Our data partner, ICES, has around **100 unique datasets**.
  → The first phase is to **pick** the relevant ones for our study.

- Some datasets are **extremely large**:
  - > 2B rows
  - > 50GB
  - Does not fit in the RAM!
    → We read the dataset by chunks of 2m rows.

- Files are in SAS (not cool for ML).

C. Data - datasets

- **Stationary** data (3 datasets)
- **Geographical** data (yearly) (4 datasets)
- **Chronic** diseases (yearly binary flags) (6 datasets)
- **Observations** (heterogenous times) (5 datasets)
- **Lab** values (1 dataset)
- The split is based on the **temporality** of the data.

19 datasets in total

layer 6

C. Data - features

Features include:

- Year of birth
- Gender

- Latitude, longitude of the patient's address

- Flag for asthma

- Physician specialty
- Diagnosis code linked to a hospitalization
- Quantity of drugs bought at a given time
- A1C measurement (important for diabetes)
- Cholesterol measurement
- ...

layer 6

C. Data - features

Features include:

- Year of birth
- Gender

Typical demographic data

- Latitude, longitude of the patient's address

neighborhood information

- Flag for asthma

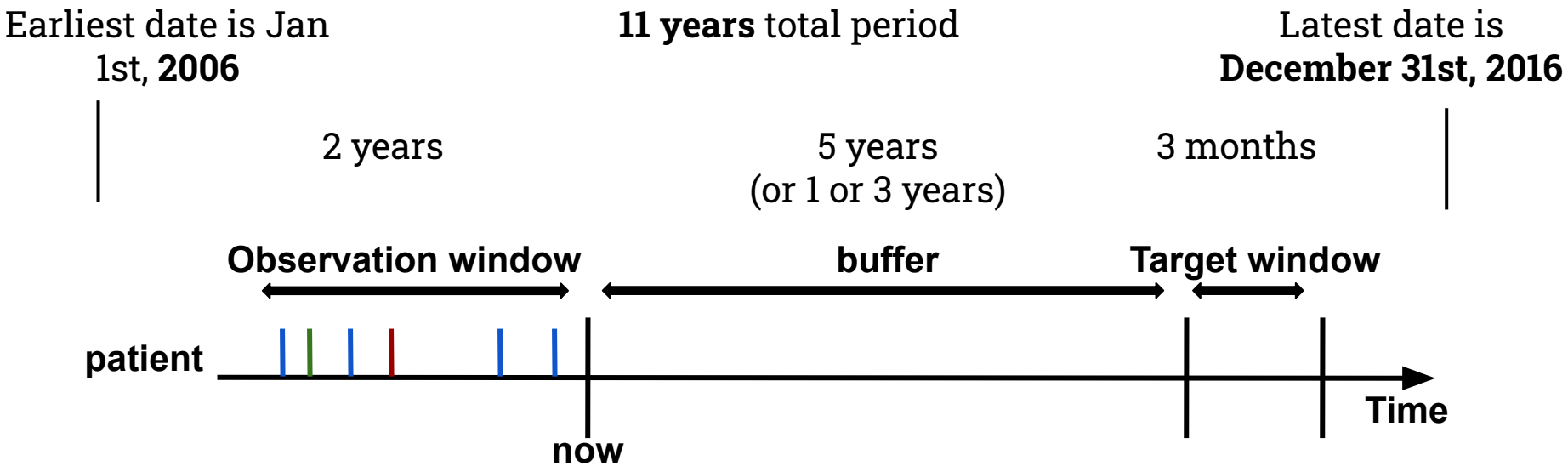
Built-in flags for chronic diseases

- Physician specialty
- Diagnosis code linked to a hospitalization
- Quantity of drugs bought at a given time
- A1C measurement (important for diabetes)
- Cholesterol measurement
- ...

Diverse observations
(codes, lab values, etc)

layer 6

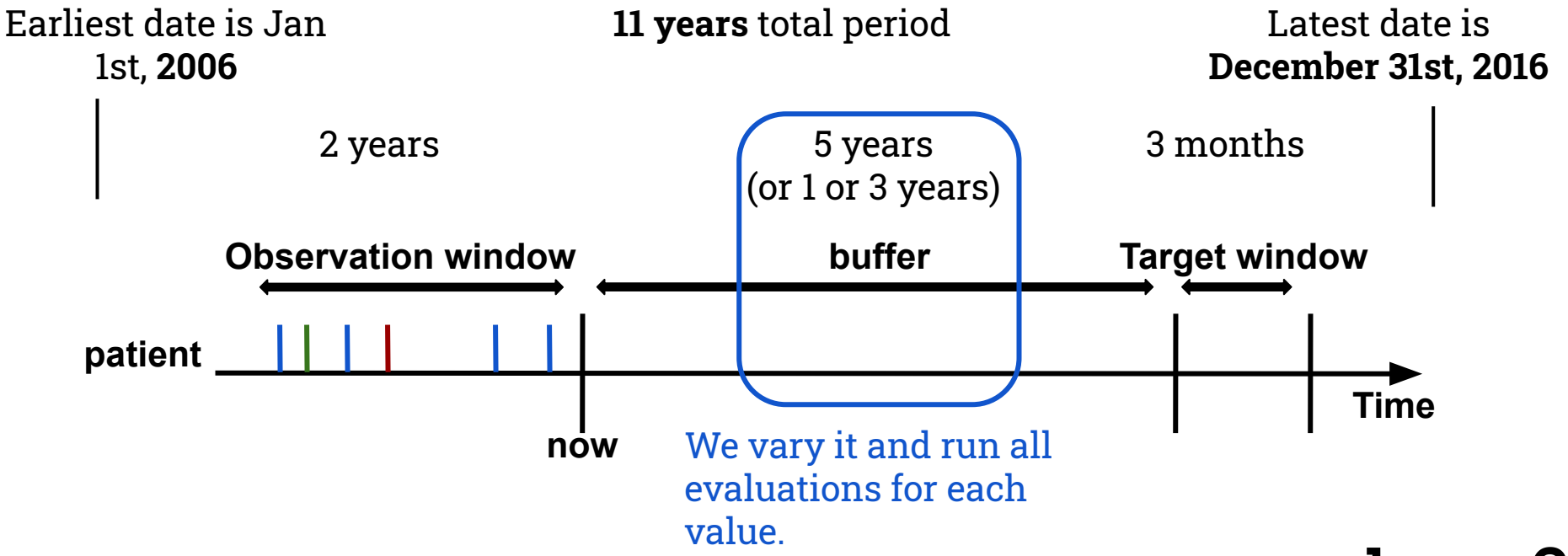
D. Framework



General way to approach time series problems...

layer 6

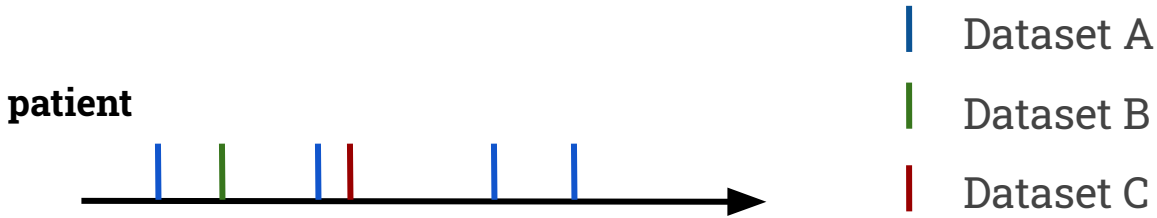
D. Framework



layer 6

D. Framework - data preparation

- Multiple, **varying numbers of observations** from each dataset in a patient's recent history:



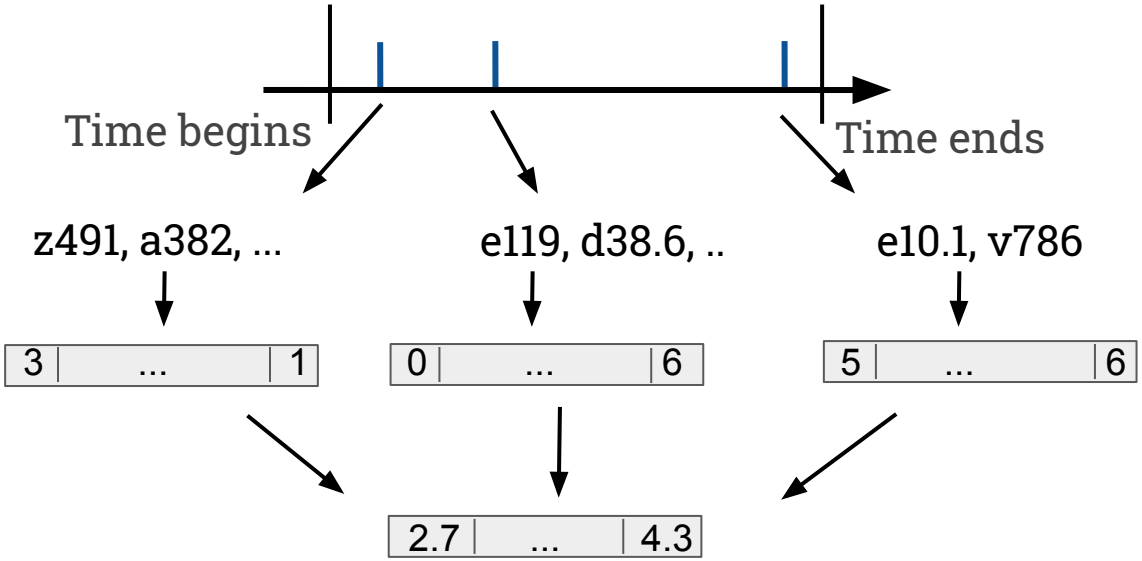
Each color maps to a unique dataset.

- We also have **stationary** patient information: birth year, etc
- How to summarize data into a **fixed-length** vector?

layer 6

D. Framework - data preparation

Let's say we have multiple observations from the same dataset in a given time window:



Averaging is just one way to aggregate...

Time bin contains 3 observations from dataset A.

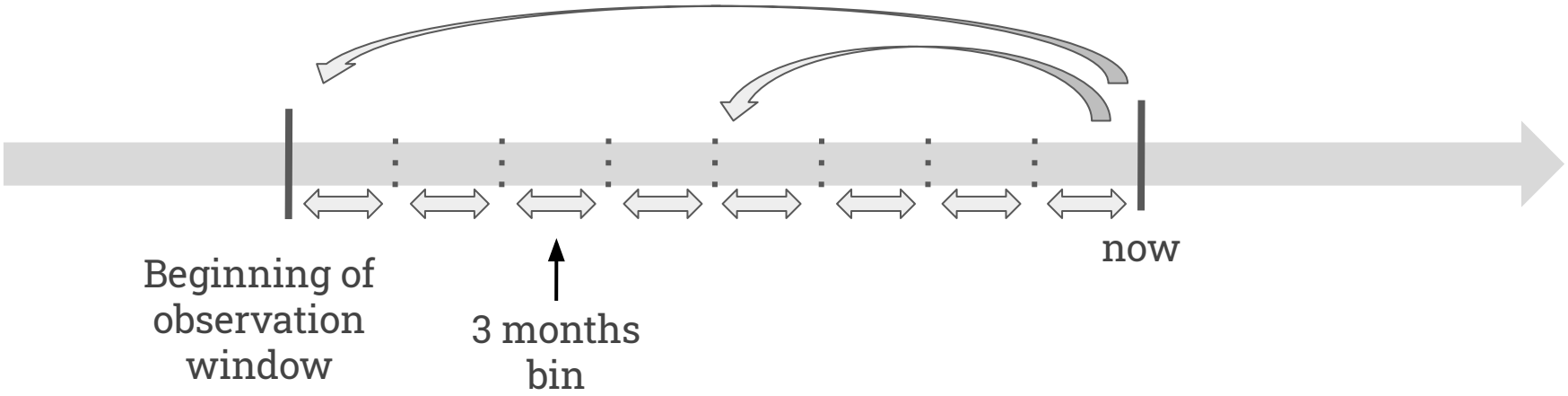
Dummy variable (1-hot or k-hot)

Averaging values gives us a **unique vector.**

layer 6

D. Framework - data preparation

There are several ways to define **aggregation periods**.
For instance when looking at 2 years of patient's history:



layer 6

D. Framework - data preparation

Averaging is not the only way to aggregate temporal data:

- Standard deviation
- Minimum, maximum
- Amplitude (max - min)
  → Needs at least two observations
- Trend (last - first)
  → Needs at least two observations
- ...

layer 6

D. Framework - data preparation

Some **manually-engineered features** turn out to be very important:

- **Age**
- Age when arriving in Canada
- **Time since** last observation from a given dataset.
- Time since last measurement of given lab.
- Mean/std of **time between** consecutive observations
- **Count** and cumulative count of observations.
- Complications **history**.
- …

Be **creative!**

D. Framework

- Our model includes several **stationary** features:
  Birth year, country of birth, etc

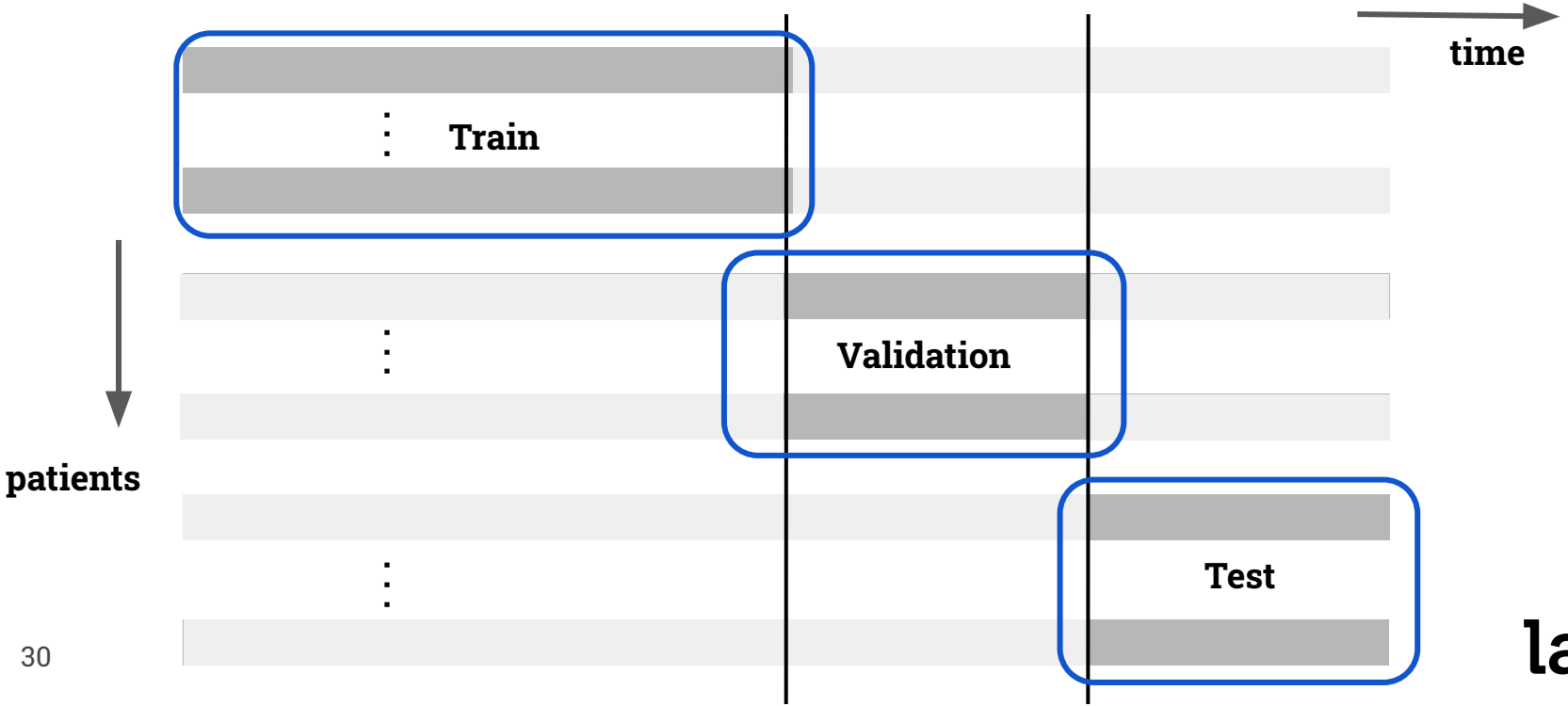
- We **concatenate** the array of stationary features to the arrays of temporal ones.

- We extract **A LOT** of features, then keep the ones with the **highest absolute feature contribution => typically top 500/700**

- No feature hand-picking.

D. Framework - split

We use a training/validation/test split in terms of **patients and time**:



layer 6

Takeaways on data + framework

- The way you prepare the data is really super important. Spend time on it!

- Feature engineering:
  Think about which features you could create that could be relevant for the problem you are tackling.

- Split:
  Use a clean train/val/test split.
  In almost all cases, you must **split on patients**.
  For temporal data, it is also best to **split on time.**

layer 6

E. Model

The model can be any machine learning model suitable for binary classification:

- Logistic regression
- Decision tree or random forest
- **Gradient boosting**:
  Xgboost, lightgbm, etc
- **Neural networks**:
  MLP, LSTM, GRU, Transformer, ODE-GRU, etc

E. Model: deep learning vs gradient boosting

A long-running competition between both approaches!

Deep learning:

- +++ Pros:
  - SOTA for **CV, NLP, Speech**
  - **No feature engineering**!
  - Can handle temporal data nicely (RNNs, Transformers)
- --- Cons:
  - Tricky to **optimize**
  - Need to preprocess and **normalize** input features
  - Newer SOTA models (e.g., Transformer-based) are large, GPU-demanding and slow to train
  - "***Black-box***"

Gradient boosting:

- +++ Pros:
  - SOTA for **tabular data**
  - No need to normalize input features
  - More robust to noisy features
  - Easier **interpretability**
- --- Cons:
  - Need feature engineering to reach best performance
  - Not designed for temporal data

layer 6

E. Model: deep learning vs gradient boosting

A long-running competition between both approaches!

Most real-life data

Deep learning:

- +++ Pros:
  - SOTA for **CV, NLP, Speech**
  - **No feature engineering**!
  - Can handle temporal data nicely (RNNs, Transformers)
- --- Cons:
  - Tricky to **optimize**
  - Need to preprocess and **normalize** input features
  - Newer SOTA models (e.g., Transformer-based) are large, GPU-demanding and slow to train
  - "***Black-box***"

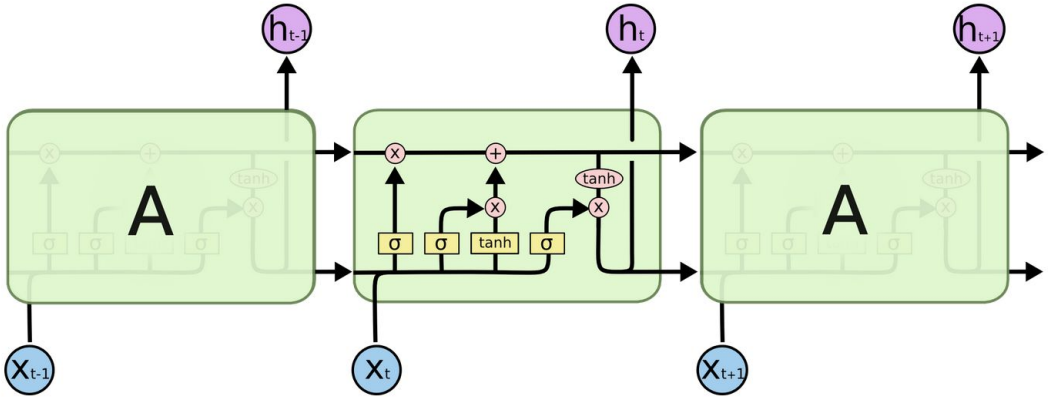
Gradient boosting:

- +++ Pros:
  - SOTA for **tabular data**
  - No need to normalize input features
  - More robust to noisy features
  - Easier **interpretability**
- --- Cons:
  - Need feature engineering to reach best performance
  - Not designed for temporal data

layer 6

E. Model: LSTMs

- "*Long Short-Term Memory*" (Hochreiter and Schmidhuber, 1997) are a variant of RNN great at modeling long-term dependency
- Brought breakthrough performance on speech (2013), then NLP (2014)

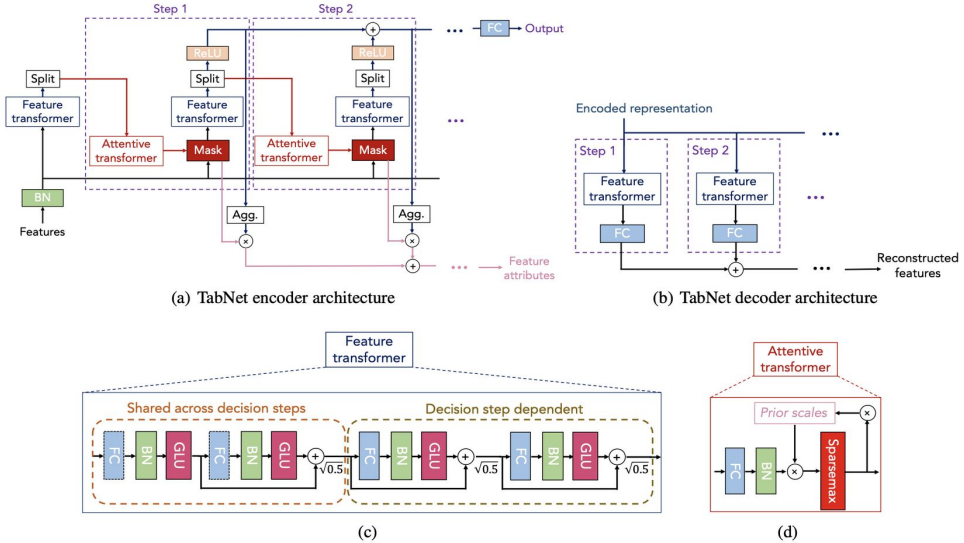


Source: *http://colah.github.io/posts/2015-08-Understanding-LSTMs/*

layer 6

E. Model: TabNet

- "*Attentive interpretable tabular learning*" (Google, 2019)
- Neural network designed for tabular data.
- Leverages a lot the attention mechanism.
- Attentive transformer decides which bits of the input features (x) it needs to pay attention (mask) at each step



(a) TabNet encoder architecture (b) TabNet decoder architecture

(c) (d)

Source: *https://sachinruk.github.io/blog/tensorflow/2021/04/05/Tabnet_From_Scratch.html*

layer 6

E. Model

The model can be any machine learning model suitable for binary classification:

- Logistic regression
- Decision tree or random forest
- **Gradient boosting**:
  Xgboost, lightgbm, etc
- **Neural networks:**
  MLP, LSTM, GRU, Transformer, ODE-GRU, etc

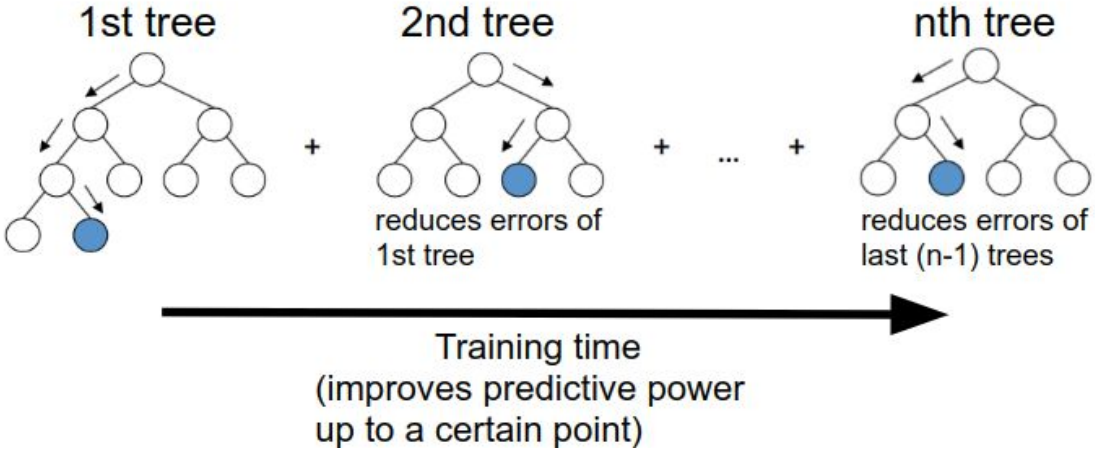
Xgboost is our favorite, because:
- Higher AUC
- Interpretable

layer 6

E. Model: xgboost

https://xgboost.readthedocs.io/en/latest/

- "X*gboost: a scalable boosting tree system*" (Tianqi Chen, 2016)
- **The unsung hero of machine learning**
- Gradient Boosting Machine with Decision Trees as Base Learners
- Use second-order Taylor expansion for gradient boosting, i.e. gradients scaled by the hessians
- Include many different regularization techniques



layer 6

E. Model: xgboost

Most important parameters to tune:

- *max_depth*
  Complexity of the base tree.
- *eta*
  Learning rate.
- *min_child_weight*
  Minimum number of instances that have to end up in each leaf.
  Increase it to fight overfitting.
- *subsample*
  Take a sample of rows.
- *colsample_bytree, colsample_bylevel, colsample_bynode*
  Take a sample of columns.

E. Model

Thoughts on choosing a model:

- A higher capacity, fancier model (ex: very deep net) is not always the best choice.

- Consider **other criteria** than performance:
  - Size and GPU need
  - Speed of training
  - **Interpretability** → **extremely important in healthcare!**
  - Processing requirements (ex: normalizing features)

- Overall, the model is less important (and less time consuming) than the data prep!

E. Model

Thoughts on choosing a model:

- A higher capacity, fancier model (ex: very deep net) is not always the best choice.

- Consider **other criteria** than performance:
  - Size and GPU need
  - Speed of training
  - **Interpretability** → **extremely important in healthcare!**
  - Processing requirements (ex: normalizing features)

- Overall, the model is less important (and less time consuming) than the data prep!

True statement

layer 6

Takeaways on the model

- Start with a **very simple model**: LR, small xgboost
  - Simple means **fast to train**.
  - Is the performance what you expected?
  - If it's **too high**: is there a **leakage**?
  - If it's **too low:** is there a **bug** in your data preparation?
  - Iterate

- Tuning parameters is less important than tuning how you prepared the data.

- Try models of **various nature**: gradient boosting, deep learning, etc.

- **Don't assume that one model will be better than all others.**
  Deep learning is only awesome on images, text or speech!

F. Results

- **Accuracy** is the most naive way to measure classification performance.
- For unbalanced tasks such as this one, the Area Under the ROC Curve (**AUC**) is better suited.
- There are **many other metrics:**
  - Precision
  - Recall
  - Lift
  - ...
- And other things to measure aside from performance!

F. Results - feature importance

- We can get the feature importance for an xgboost model with the **Shapley values**.

- Shapley values come from the **game theory** literature.
"*A Value for n-person Games*" (Shapley, Lloyd S., 1953)

- Shapley values are the **only** feature importance method satisfying a set of important properties:
  - Efficiency
  - Symmetry
  - Dummy
  - Additivity

layer 6

F. Results - calibration

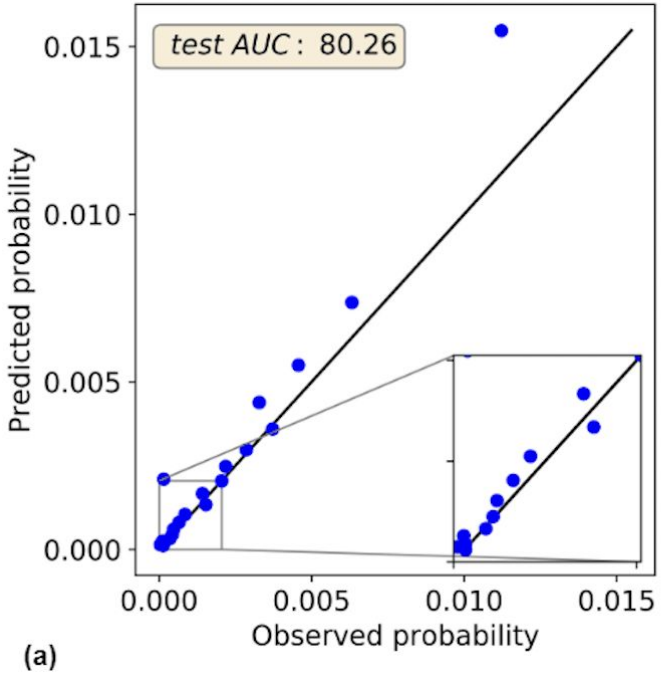
- Why bother with **calibration**?
  Our model outputs probabilities … but do these probabilities **match** with the ***actual probabilities*** of getting the complication?

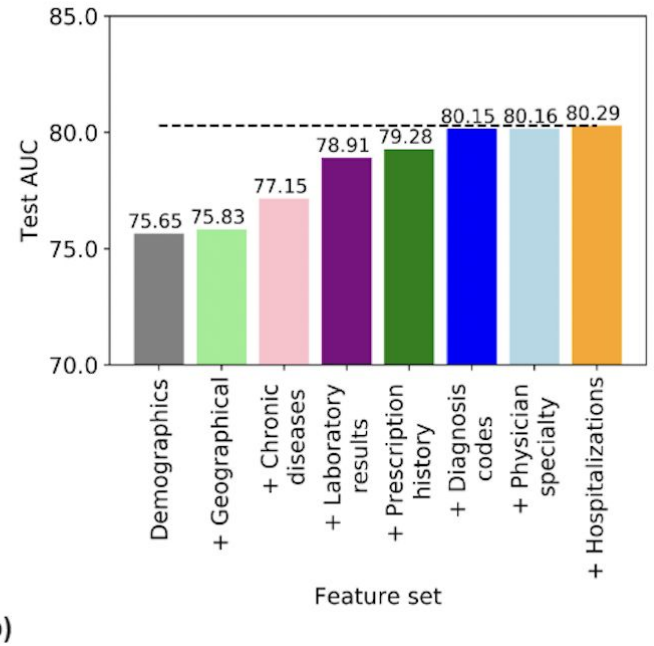
- Think about **deployment**, and a doctor reading these numbers.

- Calibration curves are simply the curves of (predicted_prob, real_prob) for bins of patients.

- We typically cut the population into 100 bins of patients and compute the mean probabilities (predicted and real) for each bin.

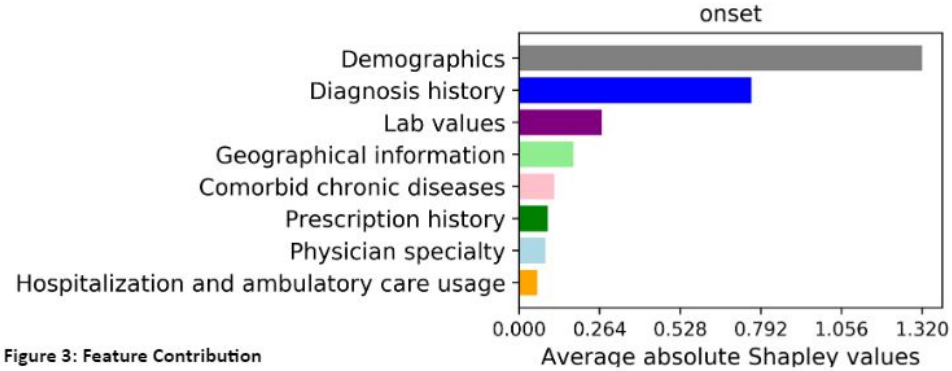
F. Results - diabetes onset (5 years ahead)





layer 6

F. Results - diabetes onset (5 years ahead)



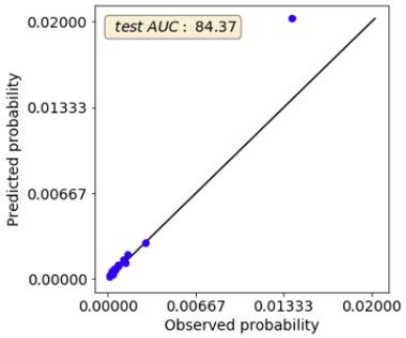
(a) Figure 3: Feature Contribution

Bin	Age	Females (%)	Immigrants (%)	Time in Canada (years)	Ethnicity marginalization score	Deprivation marginalization score	HbA1c
Model prediction							
Top 1%	58.3	59.6	38.8	17.3	4.22	3.63	5.84
Next 5%	59.4	42.3	26.5	18.4	3.85	3.45	5.81
Next 15%	58.3	40.8	16.5	19.4	3.44	3.15	5.73
Bottom 79%	31.8	55.3	11.4	19.7	3.38	2.87	5.53
Label							
Positive	53.7	49.2	19.5	19.1	3.54	3.15	5.92
Negative	37.4	52.5	13.2	19.6	3.42	2.95	5.63

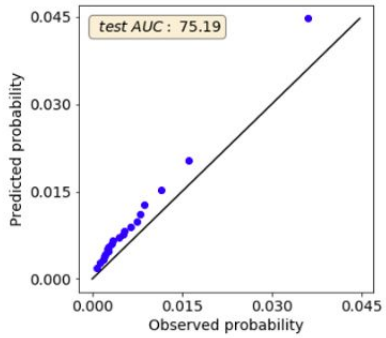
(b) Model Prediction Risk Levels

layer 6

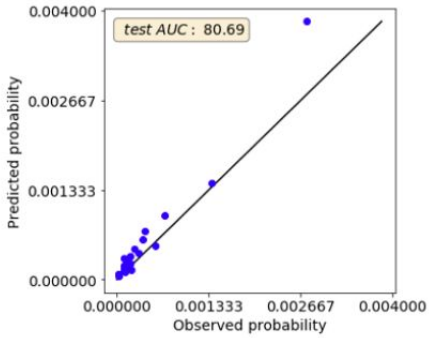
F. Results - diabetes complications (3 years)



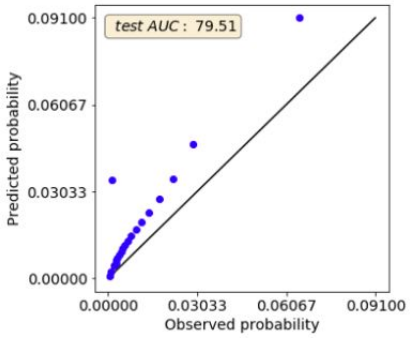
Hyper/hypo-glycemia
Test AUC range: 84.3 - 84.5



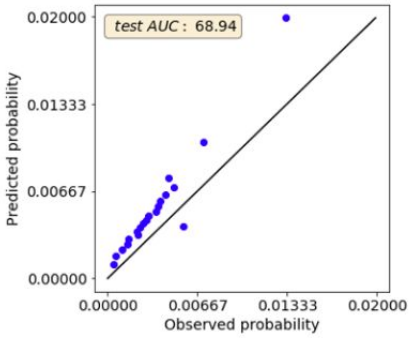
Tissue infection
Test AUC range: 75.1 - 75.2



Retinopathy
Test AUC range: 80.6 - 80.8



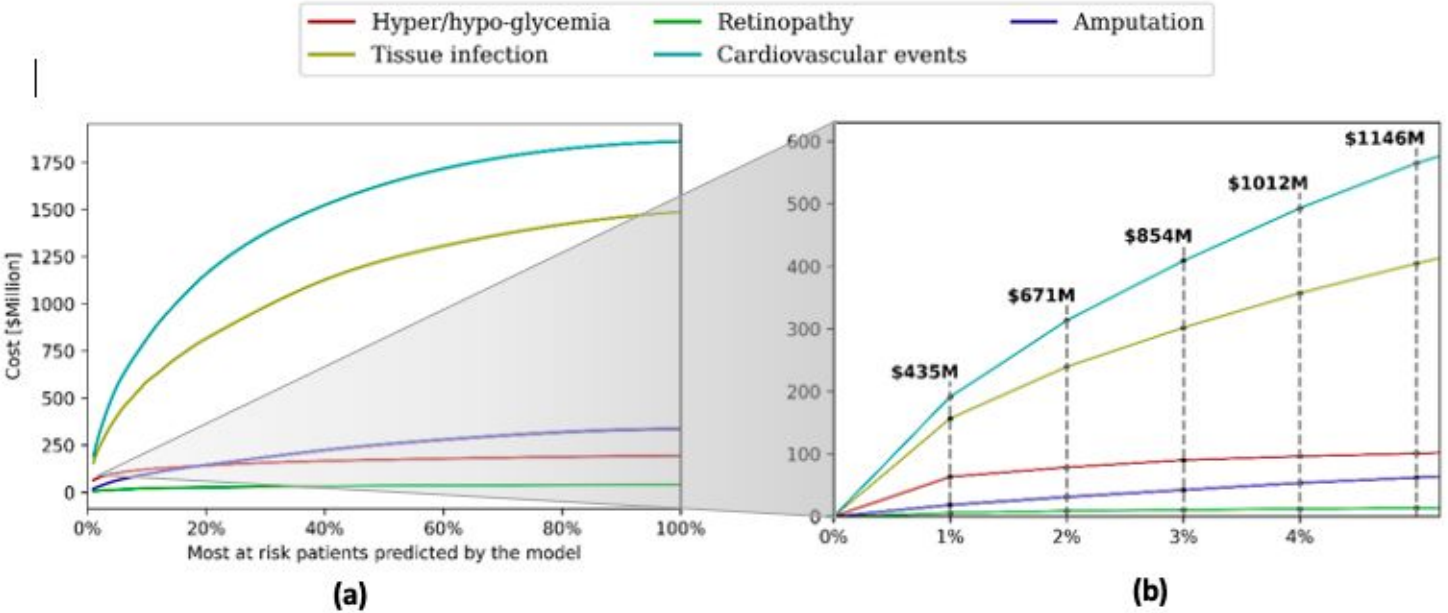
Cardiovascular events
Test AUC range: 79.4 - 79.6



Amputation
Test AUC range: 68.9 - 69.2

layer 6

F. Results - diabetes complications (3 years)



layer 6

Takeaways on the results

There are **many ways** to evaluate the performance of a machine learning model applied to a healthcare problem:

- Classification performance: AUC, (accuracy)
- Feature importance with the Shapley values
- Calibration curves
- Precision/recall/lift at different thresholds.
- ...

Ideally, if you have the time, try them all!

layer 6

G. Example

Let's apply our model to screen for **tissue infection**.

Let's look at Eddie (**fake data and name**) at the end of 2010 and at his outcome in the **last quarter of 2015**:

Feature name:	Feature importance:	Feature value:
Birth year	-0.089	1975
History of blood transfusion	-0.05	0
Quant. of medication bought in last 2 years	-0.045	0
Is the patient male?	-0.043	1
Average A1C in last 2 years	-0.04	6.1
History of diabetic complication	-0.037	0
Long-term care in the last 2 years	-0.034	0

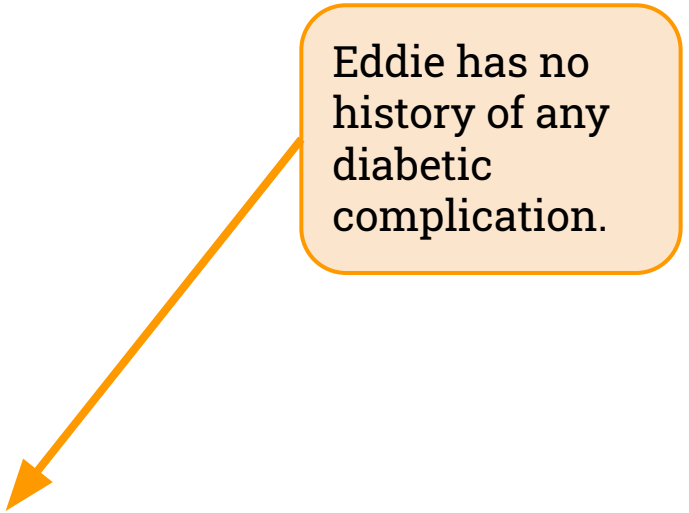
layer 6

G. Example

Let's apply our model to screen for **tissue infection**.

Let's look at Eddie (**fake data and name**) at the end of 2010 and at his outcome in the **last quarter of 2015**:

Feature name:	Feature importance:	Feature value:
Birth year	-0.089	1975
History of blood transfusion	-0.05	0
Quant. of medication bought in last 2 years	-0.045	0
Is the patient male?	-0.043	1
Average A1C in last 2 years	-0.04	6.1
History of diabetic complication	-0.037	0
Long-term care in the last 2 years	-0.034	0



layer 6

G. Example

Let's apply our model to screen for **tissue infection**.

Let's look at Eddie (**fake data and name**) at the end of 2010 and at his outcome in the **last quarter of 2015**:

Feature name:	Feature importance:	Feature value:
Birth year	-0.089	1975
History of blood transfusion	-0.05	0
Quant. of medication bought in last 2 years	-0.045	0
Is the patient male?	-0.043	1
Average A1C in last 2 years	-0.04	6.1
History of diabetic complication	-0.037	0
Long-term care in the last 2 years	-0.034	0

Which makes sense given that Eddie is young.

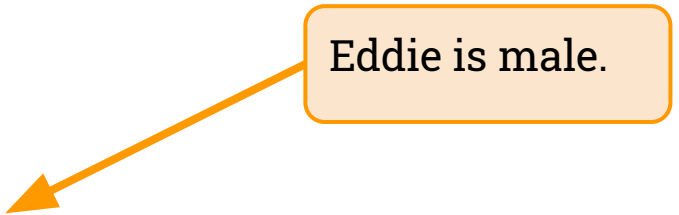
layer 6

G. Example

Let's apply our model to screen for **tissue infection**.

Let's look at Eddie (**fake data and name**) at the end of 2010 and at his outcome in the **last quarter of 2015**:

Feature name:	Feature importance:	Feature value:
Birth year	-0.089	1975
History of blood transfusion	-0.05	0
Quant. of medication bought in last 2 years	-0.045	0
Is the patient male?	-0.043	1
Average A1C in last 2 years	-0.04	6.1
History of diabetic complication	-0.037	0
Long-term care in the last 2 years	-0.034	0



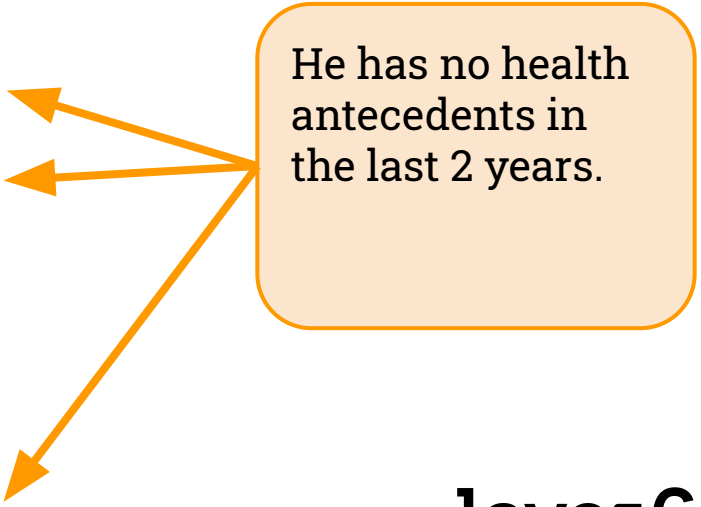
layer 6

G. Example

Let's apply our model to screen for **tissue infection**.

Let's look at Eddie (**fake data and name**) at the end of 2010 and at his outcome in the **last quarter of 2015**:

Feature name:	Feature importance:	Feature value:
Birth year	-0.089	1975
History of blood transfusion	-0.05	0
Quant. of medication bought in last 2 years	-0.045	0
Is the patient male?	-0.043	1
Average A1C in last 2 years	-0.04	6.1
History of diabetic complication	-0.037	0
Long-term care in the last 2 years	-0.034	0



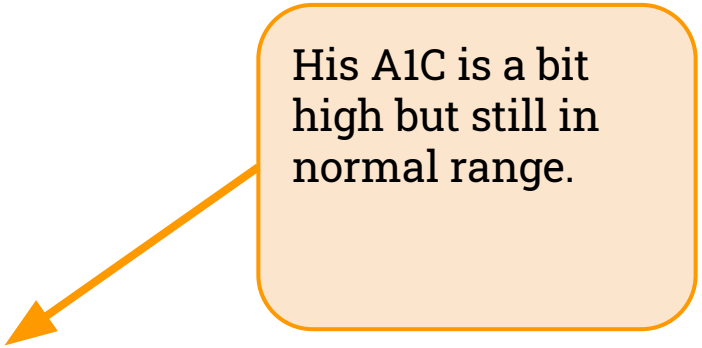
layer 6

G. Example

Let's apply our model to screen for **tissue infection**.

Let's look at Eddie (**fake data and name**) at the end of 2010 and at his outcome in the **last quarter of 2015**:

Feature name:	Feature importance:	Feature value:
Birth year	-0.089	1975
History of blood transfusion	-0.05	0
Quant. of medication bought in last 2 years	-0.045	0
Is the patient male?	-0.043	1
Average A1C in last 2 years	-0.04	6.1
History of diabetic complication	-0.037	0
Long-term care in the last 2 years	-0.034	0

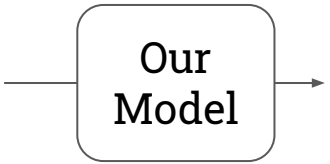


layer 6

G. Example

Given this picture of Eddie, our model predicts a **low relative value:**

Birth year	1975
History of blood transfusion	0
Quant. of medication bought in last 2 years	0
Is the patient male?	1
Average A1C in last 2 years	6.1
History of diabetic complication	0
Long-term care in the last 2 years	0



0.004 (500,000th highest)

What really happened to Eddie:

0 (no tissue infection)

layer 6

G. Example

Now let's look at another patient, Alex (**fake data and name**).
We look at Alex's history from mid-2008 to mid-2010, and at what happened to this patient in the **3rd quarter of 2015**:

Feature name:	Feature importance:	Feature value:
History of tissue infect. in ambulatory usage	0.307	4
History of diabetic complication	0.238	15
Std of time between ambulatory usage	0.091	54
Tissue infect. in ambulatory usage in last q.	0.065	3
History of tissue infect. in hospitalization	0.065	5
Diagnostic of abscess in last 2 years	0.056	1
Birth year	0.053	1952

layer 6

G. Example

Now let's look at another patient, Alex (**fake data and name**).
We look at Alex's history from mid-2008 to mid-2010, and at what happened to this patient in the **3rd quarter of 2015**:

Feature name:	Feature importance:	Feature value:
History of tissue infect. in ambulatory usage	0.307	4
History of diabetic complication	0.238	15
Std of time between ambulatory usage	0.091	54
Tissue infect. in ambulatory usage in last q.	0.065	3
History of tissue infect. in hospitalization	0.065	5
Diagnostic of abscess in last 2 years	0.056	1
Birth year	0.053	1952

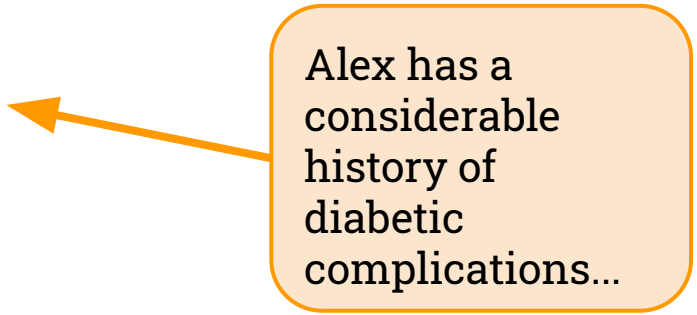
Alex is older than Chris

layer 6

G. Example

Now let's look at another patient, Alex (**fake data and name**).
We look at Alex's history from mid-2008 to mid-2010, and at what happened to this patient in the **3rd quarter of 2015**:

Feature name:	Feature importance:	Feature value:
History of tissue infect. in ambulatory usage	0.307	4
History of diabetic complication	0.238	15
Std of time between ambulatory usage	0.091	54
Tissue infect. in ambulatory usage in last q.	0.065	3
History of tissue infect. in hospitalization	0.065	5
Diagnostic of abscess in last 2 years	0.056	1
Birth year	0.053	1952

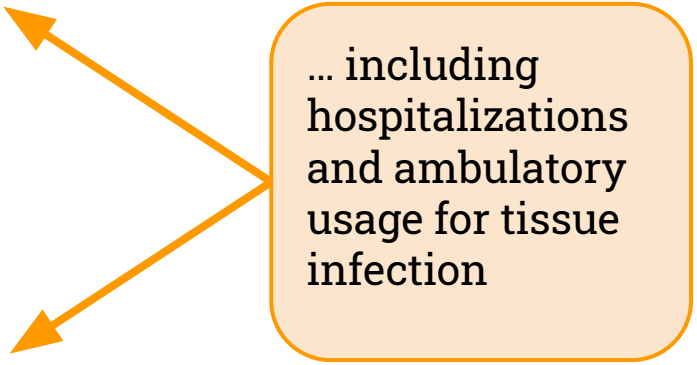


layer 6

G. Example

Now let's look at another patient, Alex (**fake data and name**).
We look at Alex's history from mid-2008 to mid-2010, and at what happened to this patient in the **3rd quarter of 2015**:

Feature name:	Feature importance:	Feature value:
History of tissue infect. in ambulatory usage	0.307	4
History of diabetic complication	0.238	15
Std of time between ambulatory usage	0.091	54
Tissue infect. in ambulatory usage in last q.	0.065	3
History of tissue infect. in hospitalization	0.065	5
Diagnostic of abscess in last 2 years	0.056	1
Birth year	0.053	1952

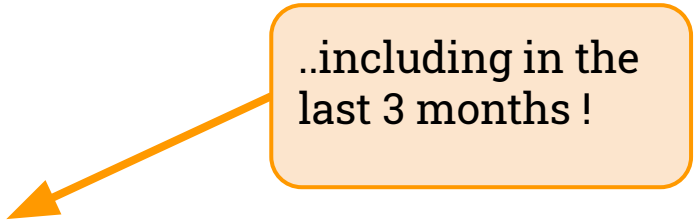


layer 6

G. Example

Now let's look at another patient, Alex (**fake data and name**).
We look at Alex's history from mid-2008 to mid-2010, and at what happened to this patient in the **3rd quarter of 2015**:

Feature name:	Feature importance:	Feature value:
History of tissue infect. in ambulatory usage	0.307	4
History of diabetic complication	0.238	15
Std of time between ambulatory usage	0.091	54
Tissue infect. in ambulatory usage in last q.	0.065	3
History of tissue infect. in hospitalization	0.065	5
Diagnostic of abscess in last 2 years	0.056	1
Birth year	0.053	1952

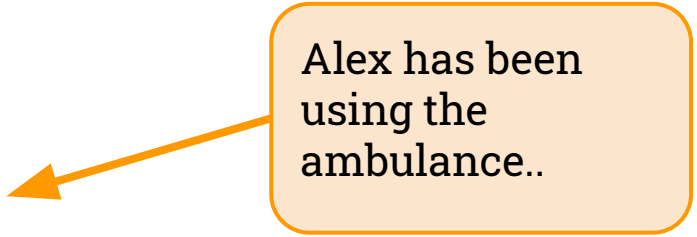


layer 6

G. Example

Now let's look at another patient, Alex (**fake data and name**).
We look at Alex's history from mid-2008 to mid-2010, and at what happened to this patient in the **3rd quarter of 2015**:

Feature name:	Feature importance:	Feature value:
History of tissue infect. in ambulatory usage	0.307	4
History of diabetic complication	0.238	15
Std of time between ambulatory usage	0.091	54
Tissue infect. in ambulatory usage in last q.	0.065	3
History of tissue infect. in hospitalization	0.065	5
Diagnostic of abscess in last 2 years	0.056	1
Birth year	0.053	1952

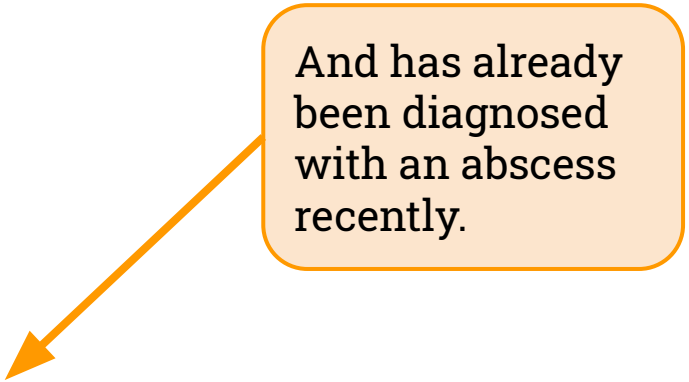


layer 6

G. Example

Now let's look at another patient, Alex (**fake data and name**).
We look at Alex's history from mid-2008 to mid-2010, and at what happened to this patient in the **3rd quarter of 2015**:

Feature name:	Feature importance:	Feature value:
History of tissue infect. in ambulatory usage	0.307	4
History of diabetic complication	0.238	15
Std of time between ambulatory usage	0.091	54
Tissue infect. in ambulatory usage in last q.	0.065	3
History of tissue infect. in hospitalization	0.065	5
Diagnostic of abscess in last 2 years	0.056	1
Birth year	0.053	1952

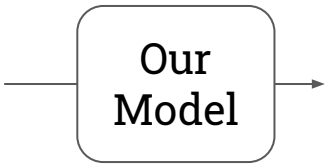


layer 6

G. Example

Given that higher age, and strong history of complications and tissue infection, our model outputs a **high risk score** to this patient:

History of tissue infect. in ambulatory usage	4
History of diabetic complication	15
Std of time between ambulatory usage	54
Tissue infect. in ambulatory usage in last q.	3
History of tissue infect. in hospitalization	5
Diagnostic of abscess in last 2 years	1
Birth year	1959



0.427

What really happened to Alex:

1 (tissue infection)

layer 6

Thank you !

Questions?

Similar applications of ML for health
PhD life vs industry life

layer 6