Development of Clinically Informing application using Recurrent Neural Network (CIReNN) based on Common Data Model

Background: Importance of Temporal Patterns in Medical Data

Patient A

	Risk factor
Age	50
Sex	M
DM (condition)	1
Metformin (drug)	1
Sulfonylurea (drug)	1
Insulin (drug)	1

Patient B

	Risk factor
Age	50
Sex	M
DM (condition)	1
Metformin (drug)	1
Sulfonylurea (drug)	1
Insulin (drug)	1

Background: Importance of Temporal Patterns in Medical Data

Patient A

	2002	2003	2004	2005	2006
Age	50	51	52	53	54
Sex	M	M	M	M	M
DM (condition)	6	5	4	4	3
Metformin (drug)	300	360	360	360	320
Sulfonylurea (drug)	0	0	0	0	1
Insulin (drug)	0	0	0	0	1

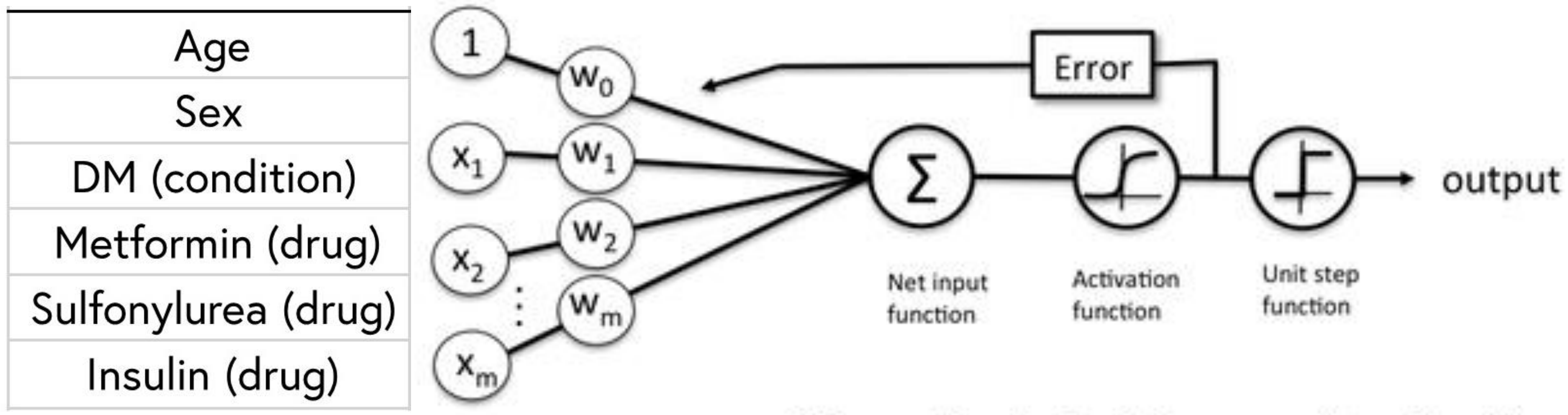
Risk factor
50
M
1
1
1
1

Patient B

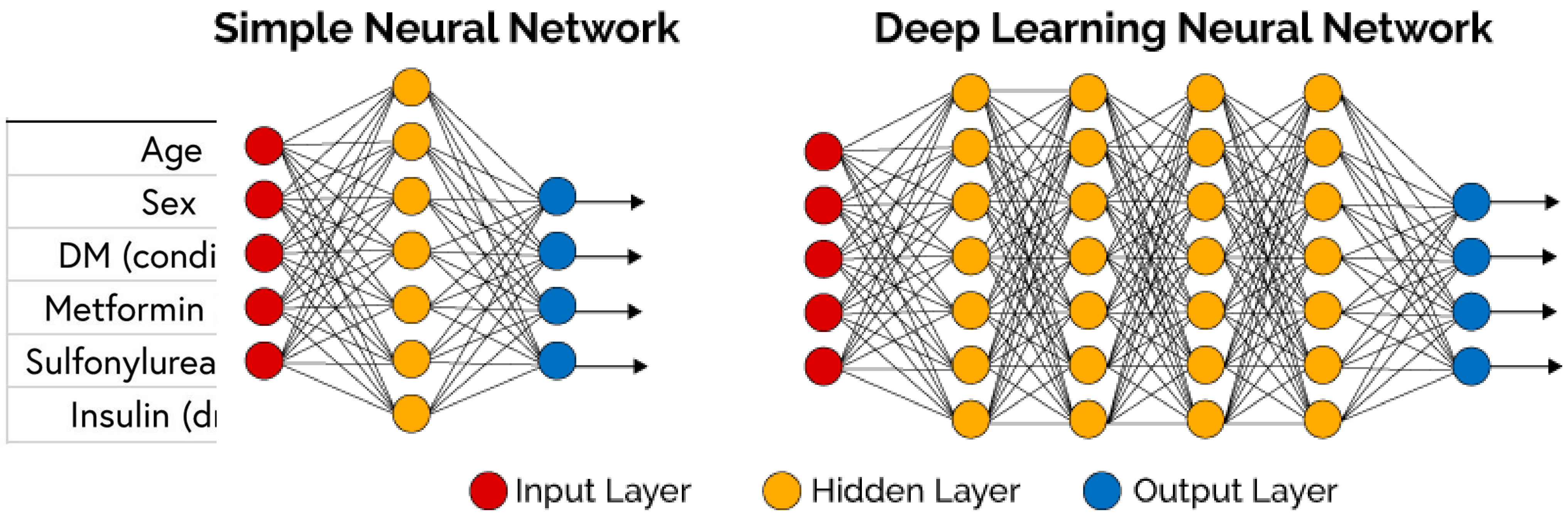
	2002	2003	2004	2005	2006
Age	50	51	52	53	54
Sex	M	M	M	M	M
DM (condition)	7	1	2	7	15
Metformin (drug)	360	10	60	150	0
Sulfonylurea (drug)	0	0	90	120	0
Insulin (drug)	0	0	0	0	360

Risk factor
50
M
1
1
1
1

Background: Deep Learning for EHR

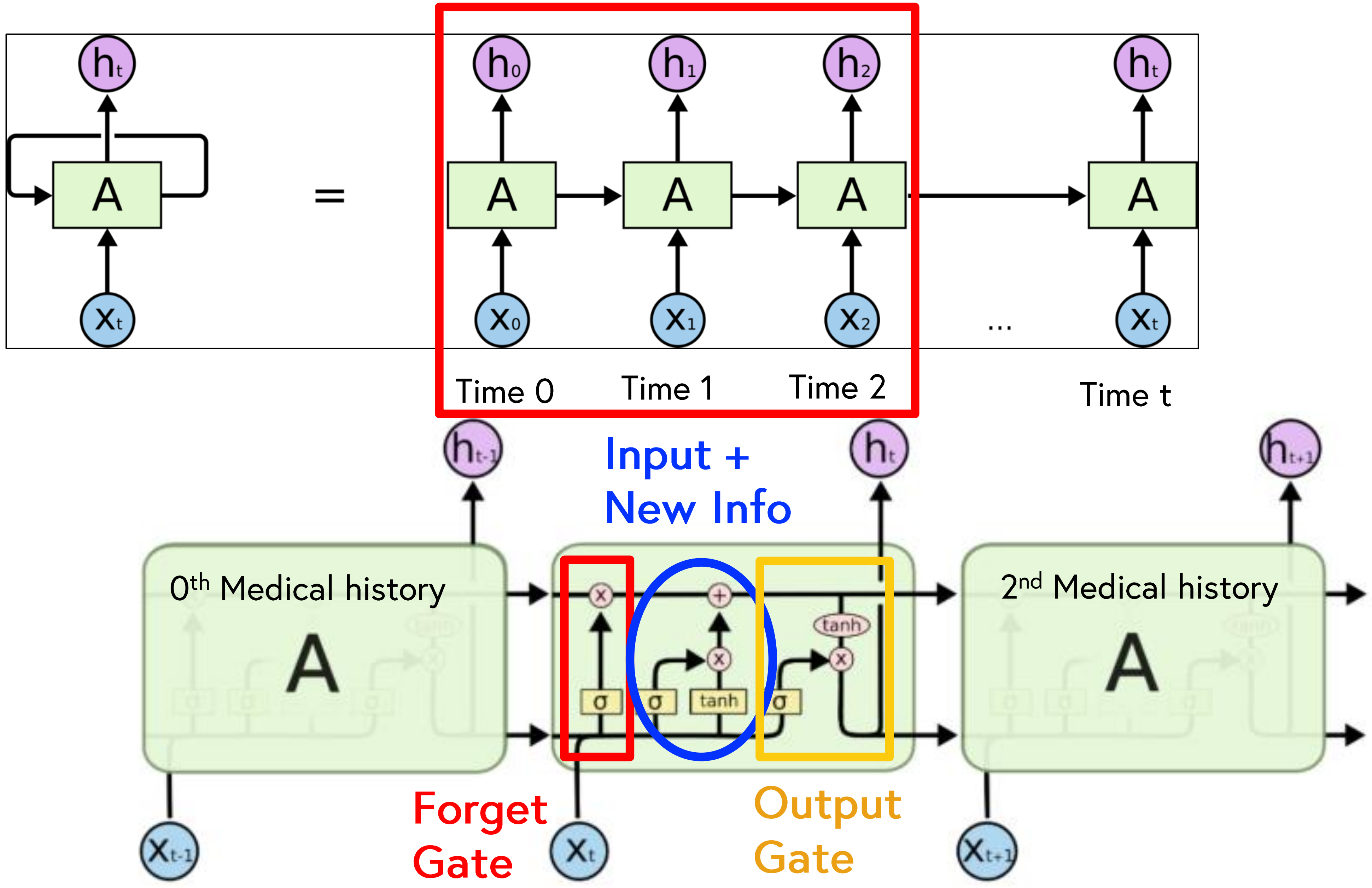


Schematic of a logistic regression classifier.



Vázquez F. Deep Learning made easy with Deep Cognition. 2017. https://becominghuman.ai/deep-learning-made-easy-with-deep-cognition-403fbe445351.

Background: Recurrent Neural Network



https://medium.com/@jianqiangma/all-about-recurrent-neural-networks-9e5ae2936f6e

Background: Deep Learning for EHR and Challenges

npj | Digital Medicine www.nature.com/npjdigitalmed

ARTICLE OPEN

Scalable and accurate deep learning with electronic health records

- It is widely held that 80% of the effort in an analytic model is preprocessing, merging, customizing, and cleaning datasets, not analyzing them for insights. **This profoundly limits the scalability of predictive models**
- The number of potential predictor variables in the electronic health record (EHR) may easily number in the thousands

Rajkomar et al., Npj Digital Medicine, 2018

Purpose

- To build a predictive model based on a recurrent neural network by using temporal features extracted from **OMOP-CDM** database: **ClReNN** (Clinically Informing application based on Recurrent Neural Network)
- **ClReNN** is expected to facilitate prediction of important clinical events by analyzing flexible and temporal relationships in health care data

Method: Data Representation

- User-defined time span based data representation
  - Regular number of time step
  - Represent the length of time and acuity/chronicity
  - Can be applied to medical data with both long-term and short-term observation period

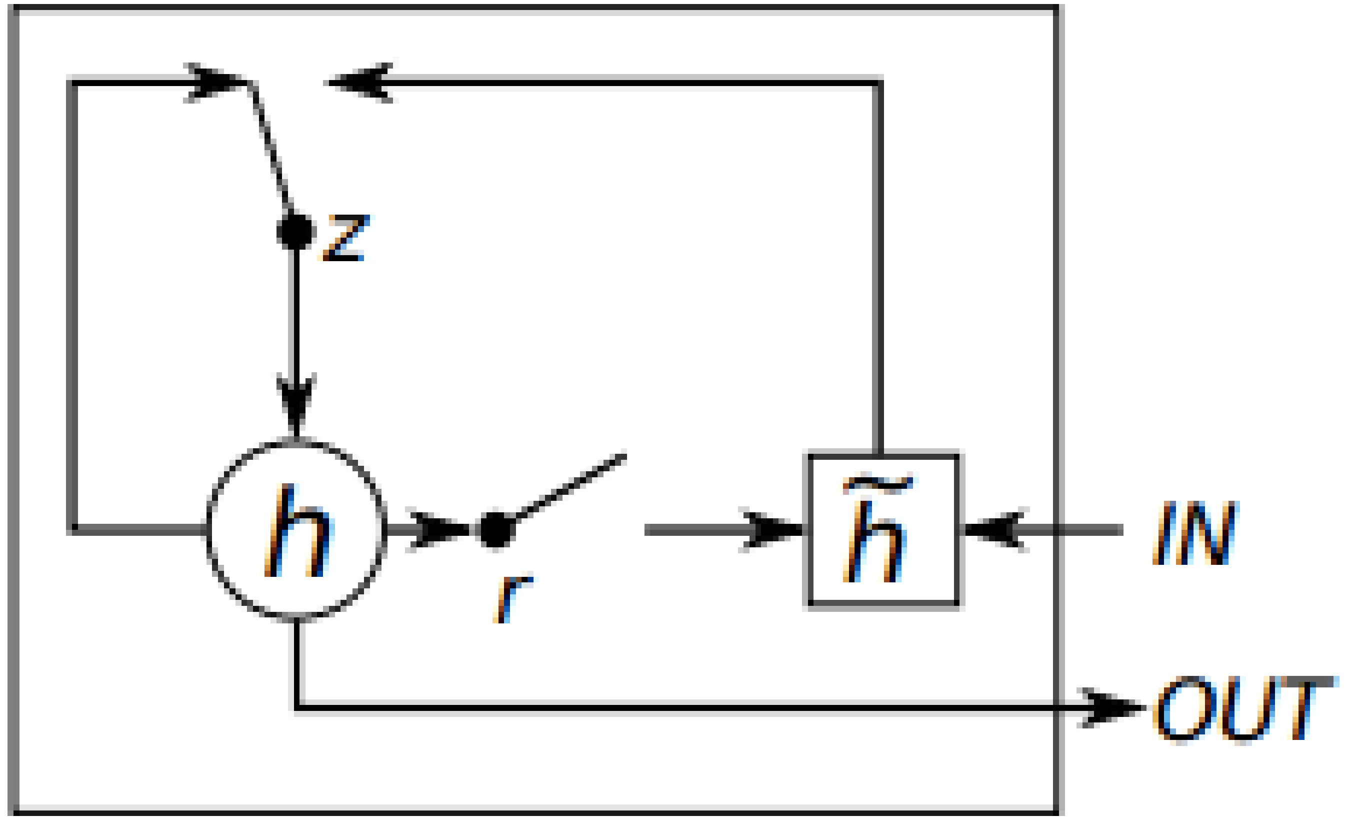
N-dimensional vector

N-dimensional vectors (**sparse array**)

Time span 1 (-1800~1770 day): Fever, Cough[condition], Tylenol[drug]	[1,1,0,0,0,…0,0,1,0]
Time span 2 (-1770~1740 day): Pneumonia [condition], Tylenol, Amoxicillin[drug]	[0,0,1,0,0,…0,0,1,1]
…	…
Time span 50 (-30~ index day): Diabetes mellitus [condition], Insulin[drug]	[0,0,0,1,0,…0,0,1,1]

Method: Deep Learning Model

- GRU layer
    - 1 layer
    - Keras with tensorflow backend in R
- Prediction layer
    - Optimizer: RMSprop
    - Loss: Cross-entropy
- For efficiency
    - Sparse array
    - Batch generator
    - Early stopping based on loss of validation set
- Sensitivity vs Specificity
    - Outcome weight
- Greedy search algorithm with n-fold cross-validation for optimal hyper-parameter



(b) Gated Recurrent Unit

Chung et al., arXiv, 2016

Method: Whole Process of Prediction in OHDSI Ecosystem

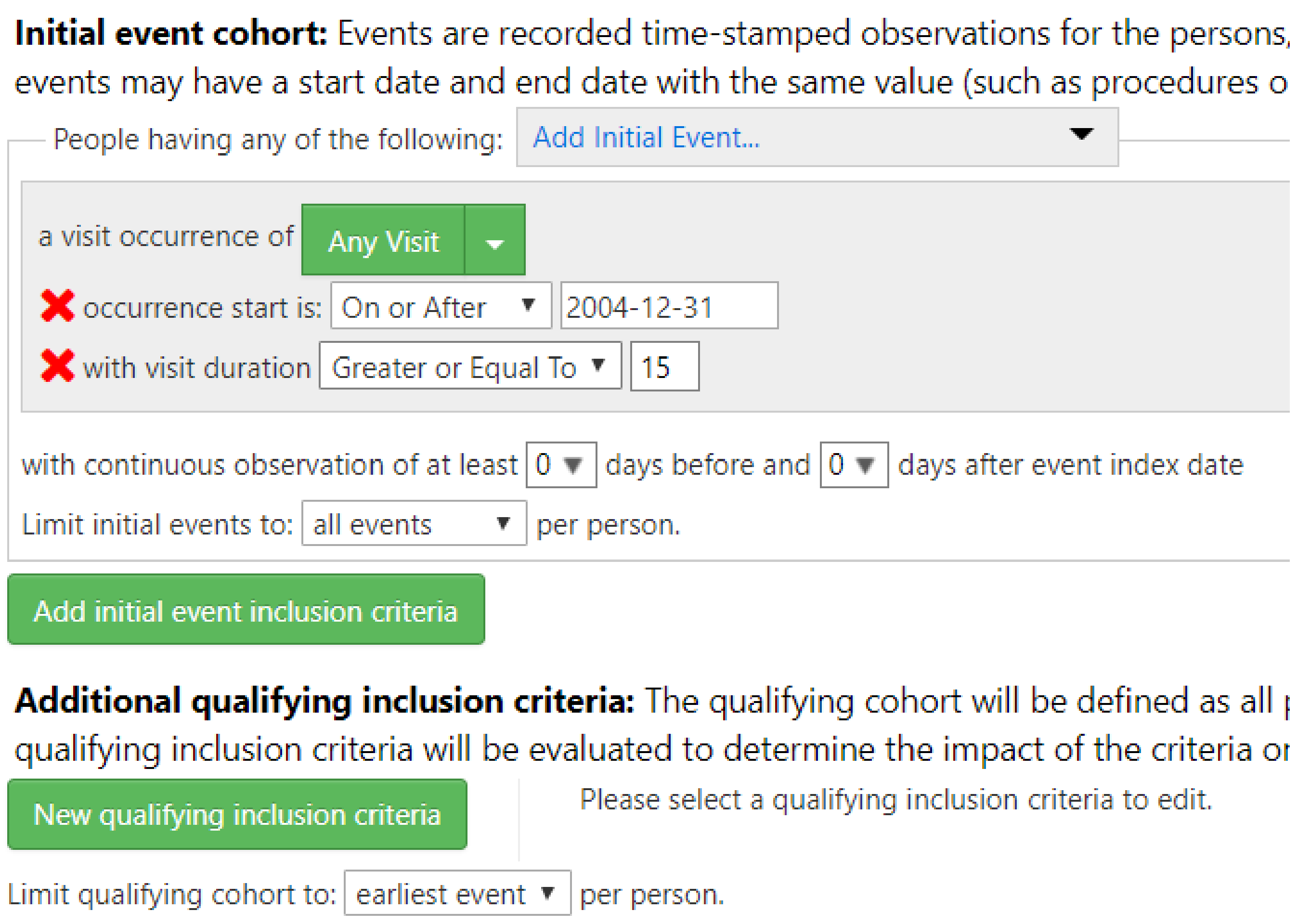
Create Target & Outcome Cohort	ATLAS
Database Connections	DatabaseConnector
Select and Extract Temporal Predictor Variables	FeatureExtraction
Develop RNN model & internal validation	PatientLevelPrediction CIReNN
Model evaluation and internal validation	PatientLevelPrediction
External validation	OHDSI network

Experiment 2: Two-week Mortality in Hospital

Create
Target & Outcome Cohort

ATLAS

- Target cohort
  - Inpatient from 1st Jan 2005
  - At least 15 days of hospitalized
- Outcome cohort
  - All-cause mortality within 14 days



http://www.ohdsi.org/web/atlas/

Experiment 2: Two-week Mortality in Hospital

Select and Extract Temporal Predictor Variables → FeatureExtraction

```
createTemporalCovariateSettings(useDemographicsGender = TRUE,
                                useDemographicsAge = TRUE,
                                useDemographicsRace = TRUE,
                                useDemographicsIndexYear = TRUE,
                                useDemographicsIndexMonth = TRUE,
                                useConditionOccurrence = TRUE,
                                useDrugExposure = TRUE,
                                useDrugEraGroupStart = TRUE,
                                useProcedureOccurrence = TRUE,
                                useMeasurement = TRUE,
                                useMeasurementValue = TRUE,
                                useVisitCount = TRUE,
                                temporalStartDays = startDays,
                                temporalEndDays = endDays)
```

- Time span
  - Risk window: 3 years + 14 days
  - Time span interval : 3-year + 2 days

https://github.com/OHDSI/FeatureExtraction

Experiment 2: Two-week Mortality in Hospital

Develop RNN model & internal validation

PatientLevelPrediction
CIReNN

```
setCIReNN(units=c(64), recurrentDropout=c(0.3),
          layerDropout=c(0.3),
          lr =c(1e-4), decay=c(1e-5),
          outcomeWeight = c(4.0),
          batchSize = c(100),
          epochs= c(100), seed=NULL)
```

```
runPlp(population,
       plpData,
       minCovariateFraction = 0.005,
       modelSettings = CIReNNModel,
       testSplit = "person",
       testFraction = 0.2,
       nfold = 5,
       save = dataFolder)
```

https://github.com/OHDSI/patientlevelprediction

Experiment 2: Two-week Mortality in Hospital

Model evaluation and internal validation → PatientLevelPrediction

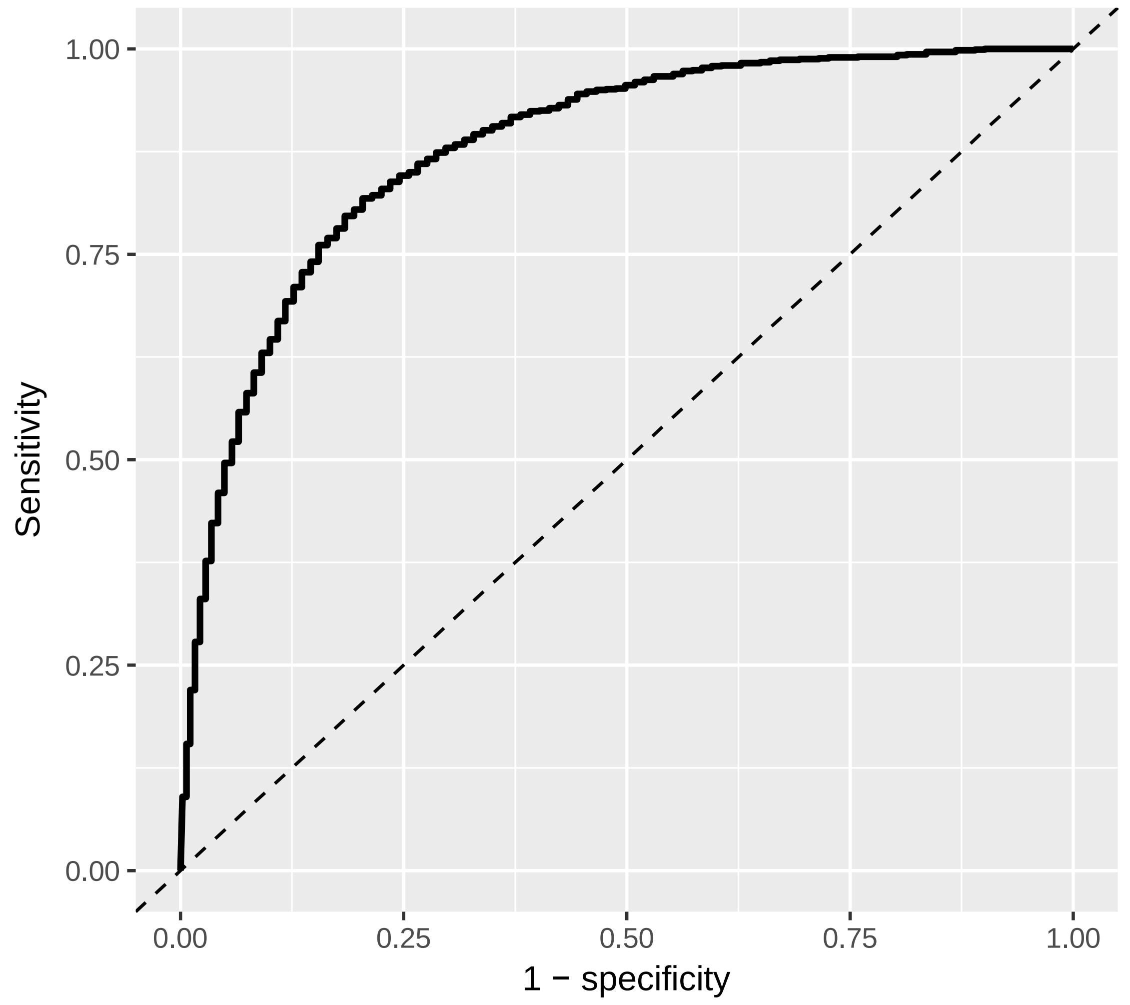
- Target population: 56,811
- Outcome cases : 5,189 (9.1%)
- 20% test and 80% train, 5-fold cross-validation
- AUROC
  - In train: 0.92
  - In test: 0.88

Experiment 2: Two-week Mortality in Hospital

Model evaluation and internal validation

PatientLevelPrediction

```
PatientLevelPrediction::plotPlp(CIReNNresult,file.path(dataFolder,"CIReNN_plot.pdf"))
```

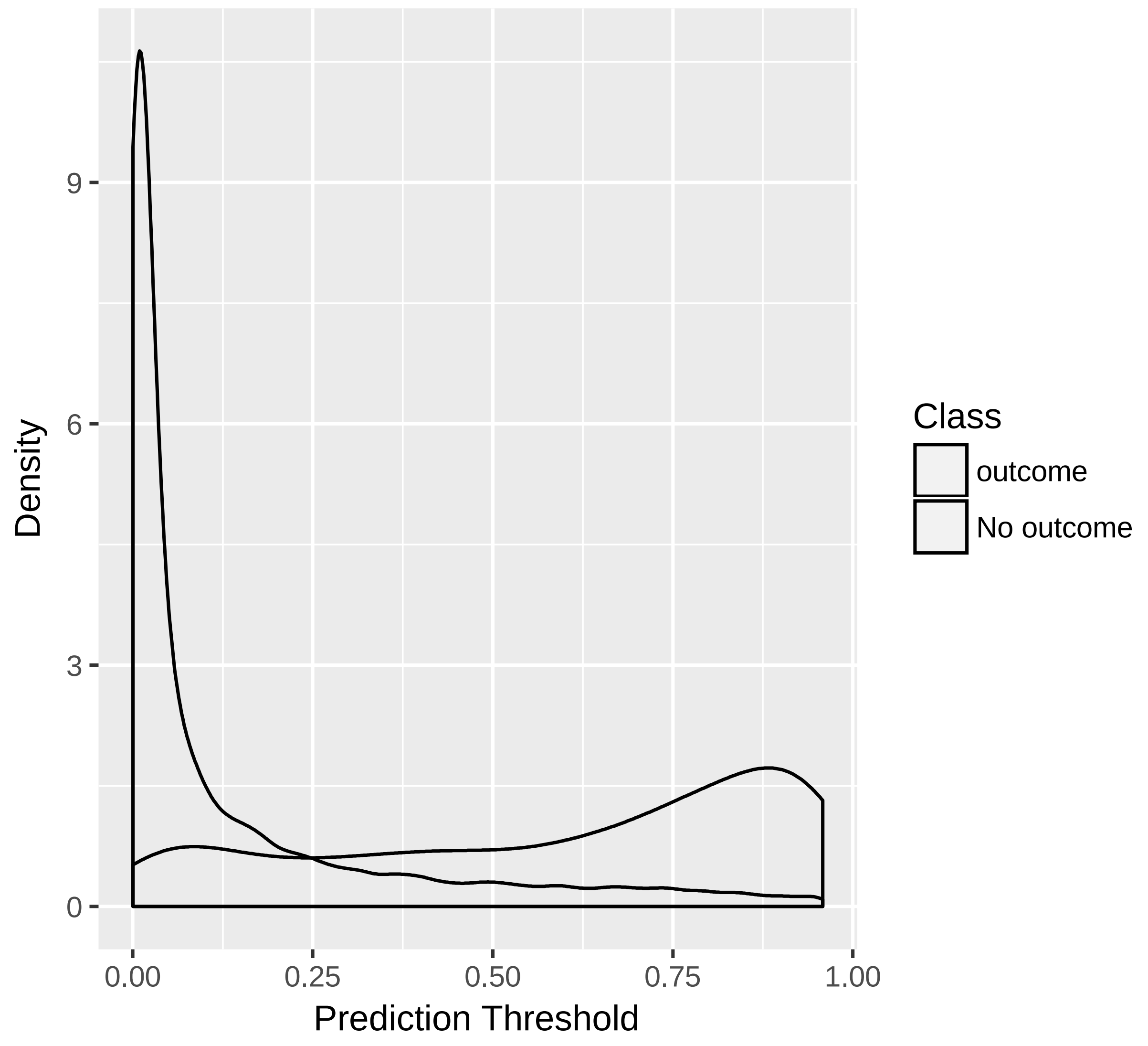


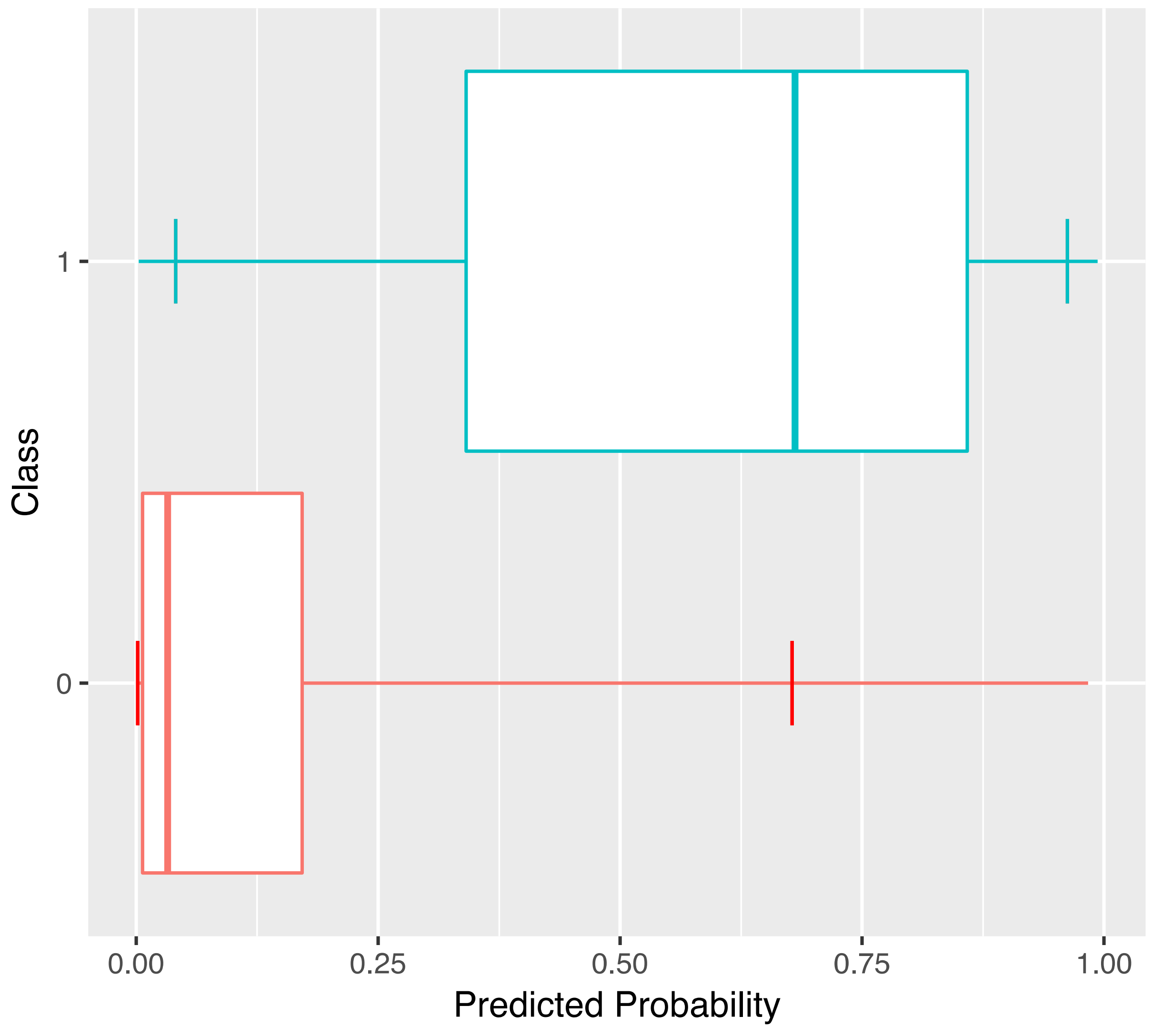
AUROC: 0.88

Experiment 2: Two-week Mortality in Hospital

Model evaluation and internal validation

PatientLevelPrediction

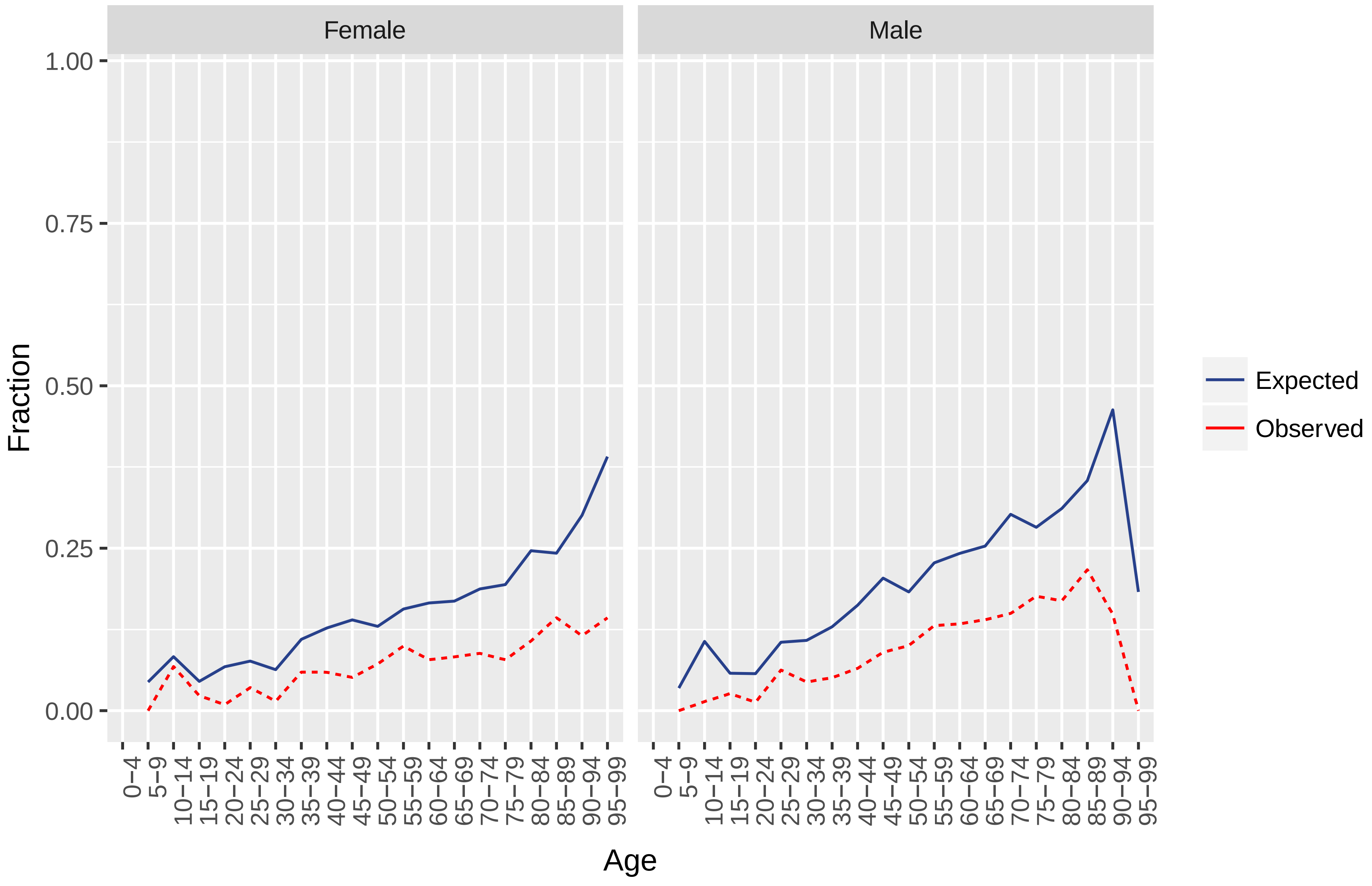




Experiment 2: Two-week Mortality in Hospital

Model evaluation and internal validation

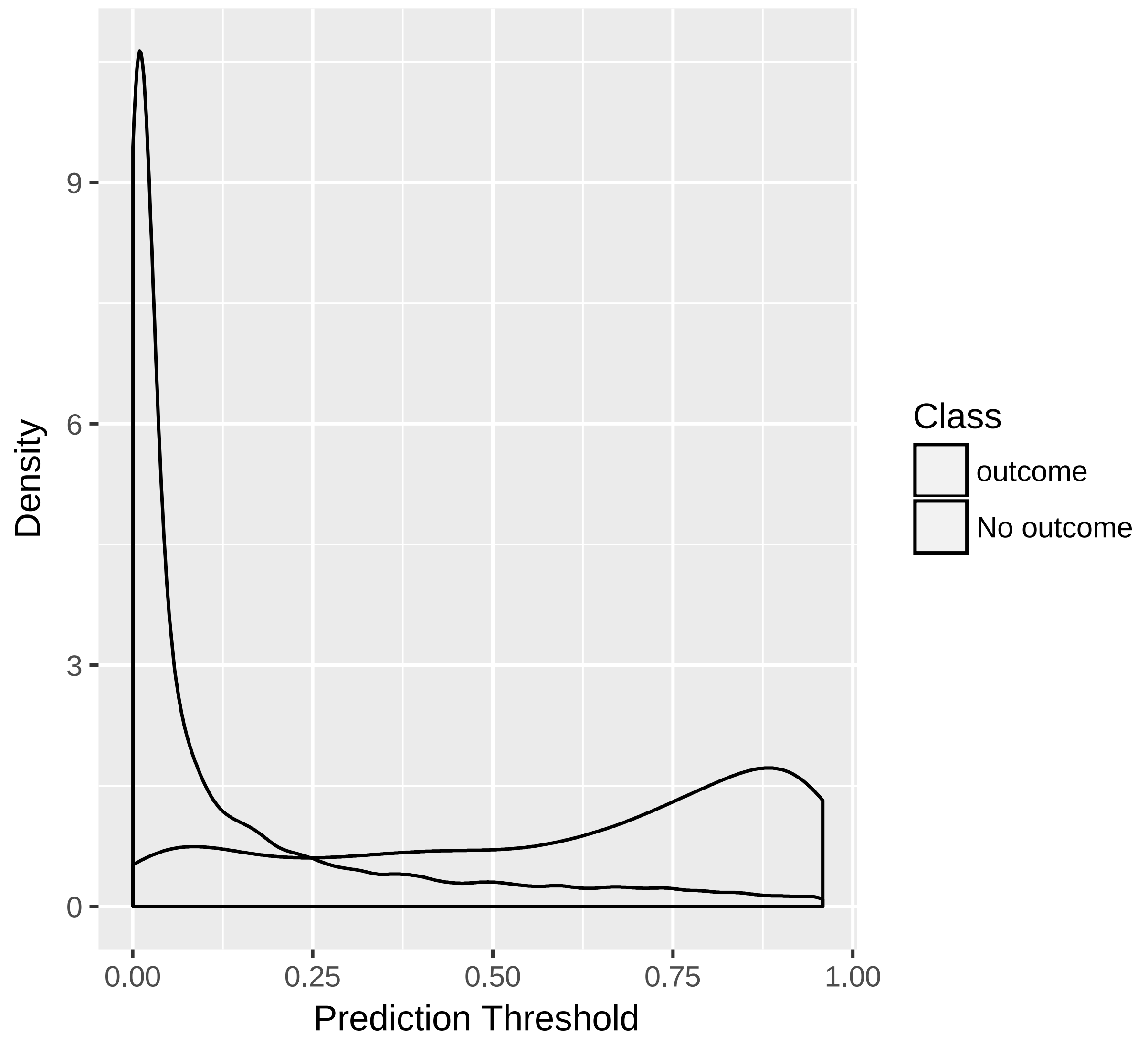
PatientLevelPrediction

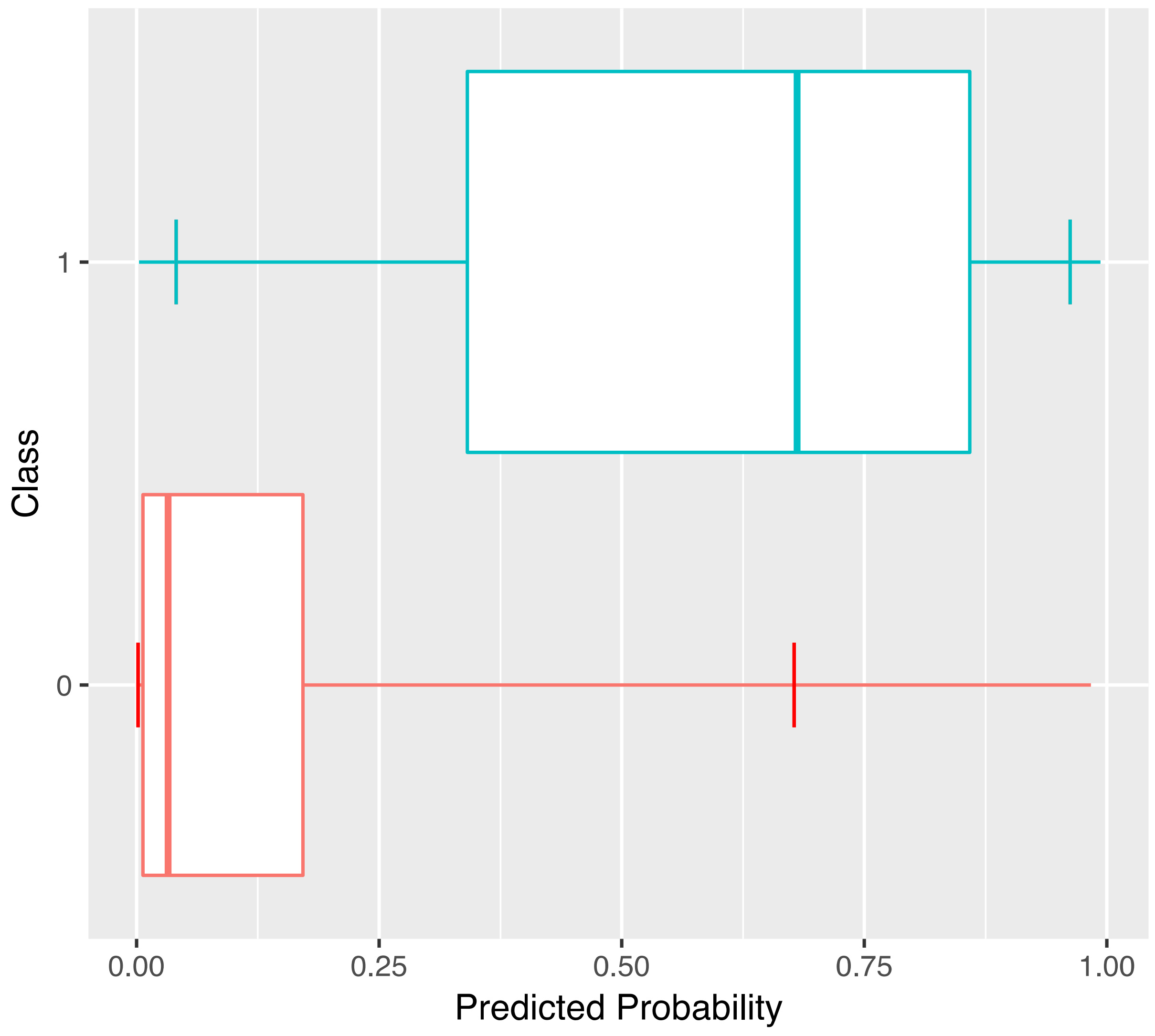


Experiment 2: Two-week Mortality in Hospital

Model evaluation and internal validation

PatientLevelPrediction

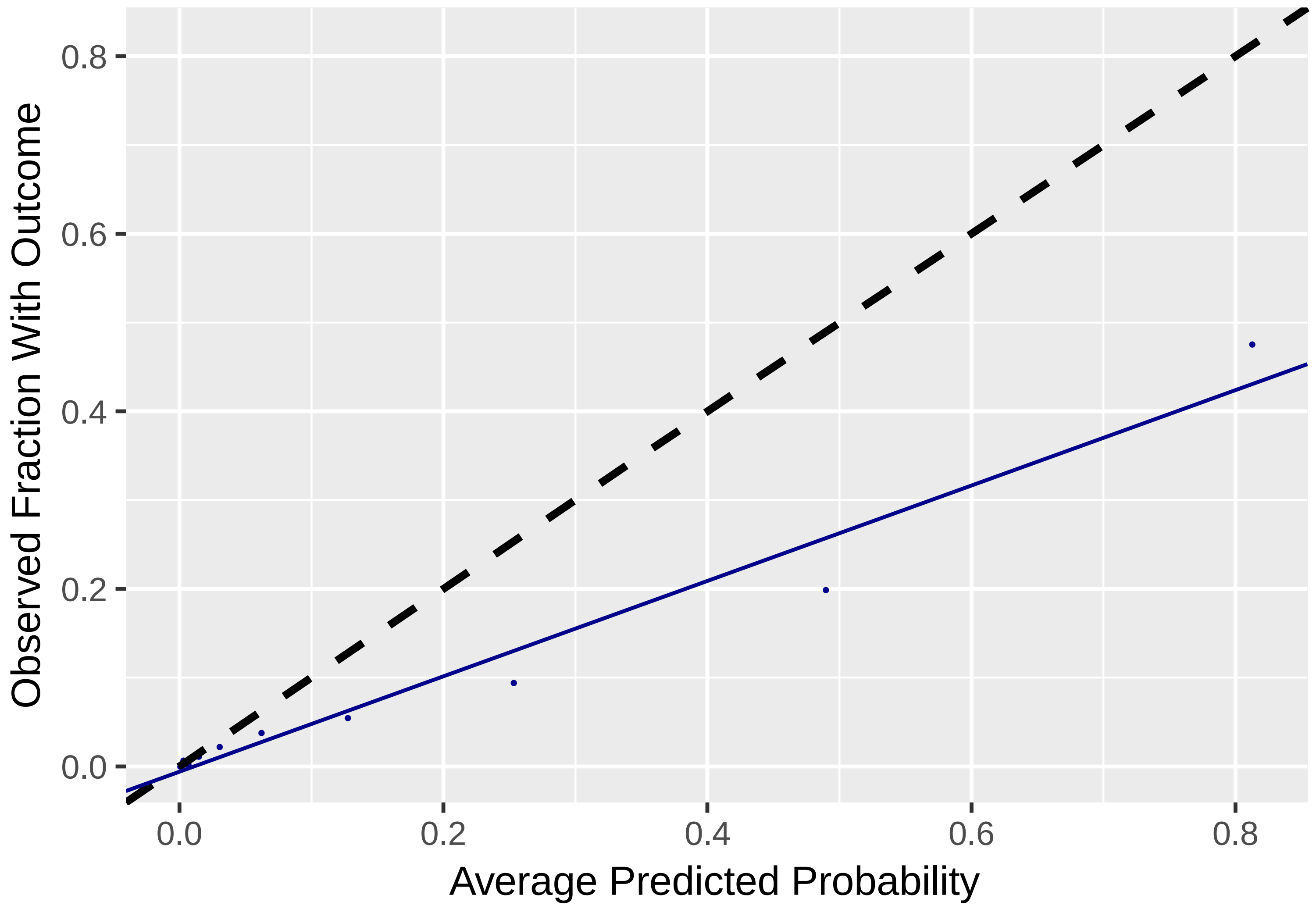




Experiment 2: Two-week Mortality in Hospital

Model evaluation and internal validation

PatientLevelPrediction



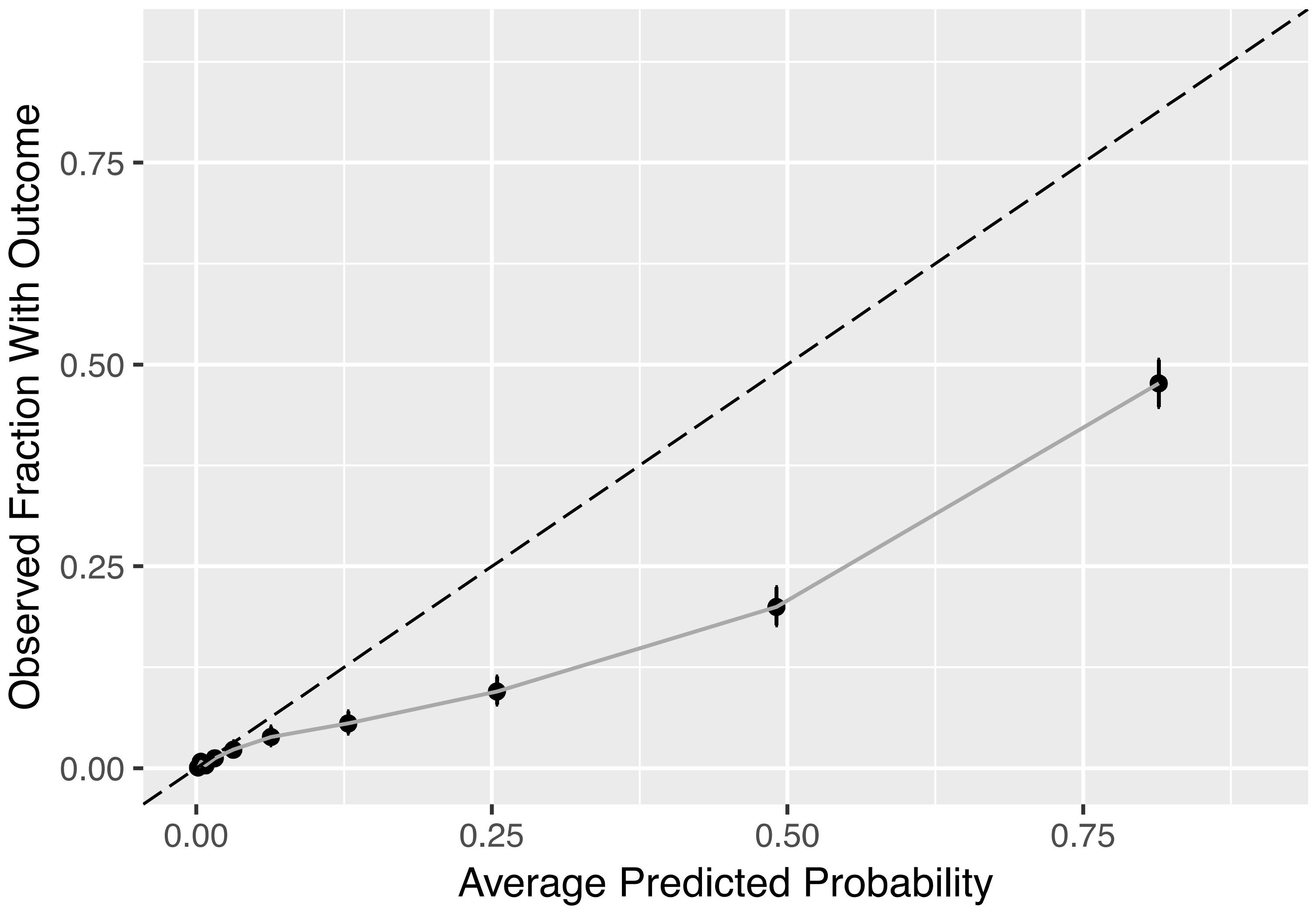


Image Sources

- Vázquez F. Deep Learning made easy with Deep Cognition. 2017. https://becominghuman.ai/deep-learning-made-easy-with-deep-cognition-403fbe445351.
- Ma J. All of Recurrent Neural Networks [Internet]. Medium. 2016. Available from: https://medium.com/@jianqiangma/all-about-recurrent-neural-networks-9e5ae2936f6e

References

- Choi E, Schuetz A, Stewart W, Sun J. Using recurrent neural network models for early detection of heart failure onset. Journal of the American Medical Informatics Association. 2017;:ocw112.
- Rajkomar A, Oren E, Chen K, Dai A, Hajaj N, Hardt M et al. Scalable and accurate deep learning with electronic health records. npj Digital Medicine. 2018;1(1).
- Chung, J., Ahn, S., & Bengio, Y. (2016). Hierarchical multiscale recurrent neural networks. arXiv preprint arXiv:1609.01704.