

Motivation

Framingham risk score (FRS), the most widely used cardiac risk scoring system, suffers from limited accuracy due to not only the small number of patient variables but also the linear statistical scheme used in FRS.

Previously, we showed that random forest (RF) non-linear machine learning with many additional clinical variables has better performance at predicting mortality than FRS.

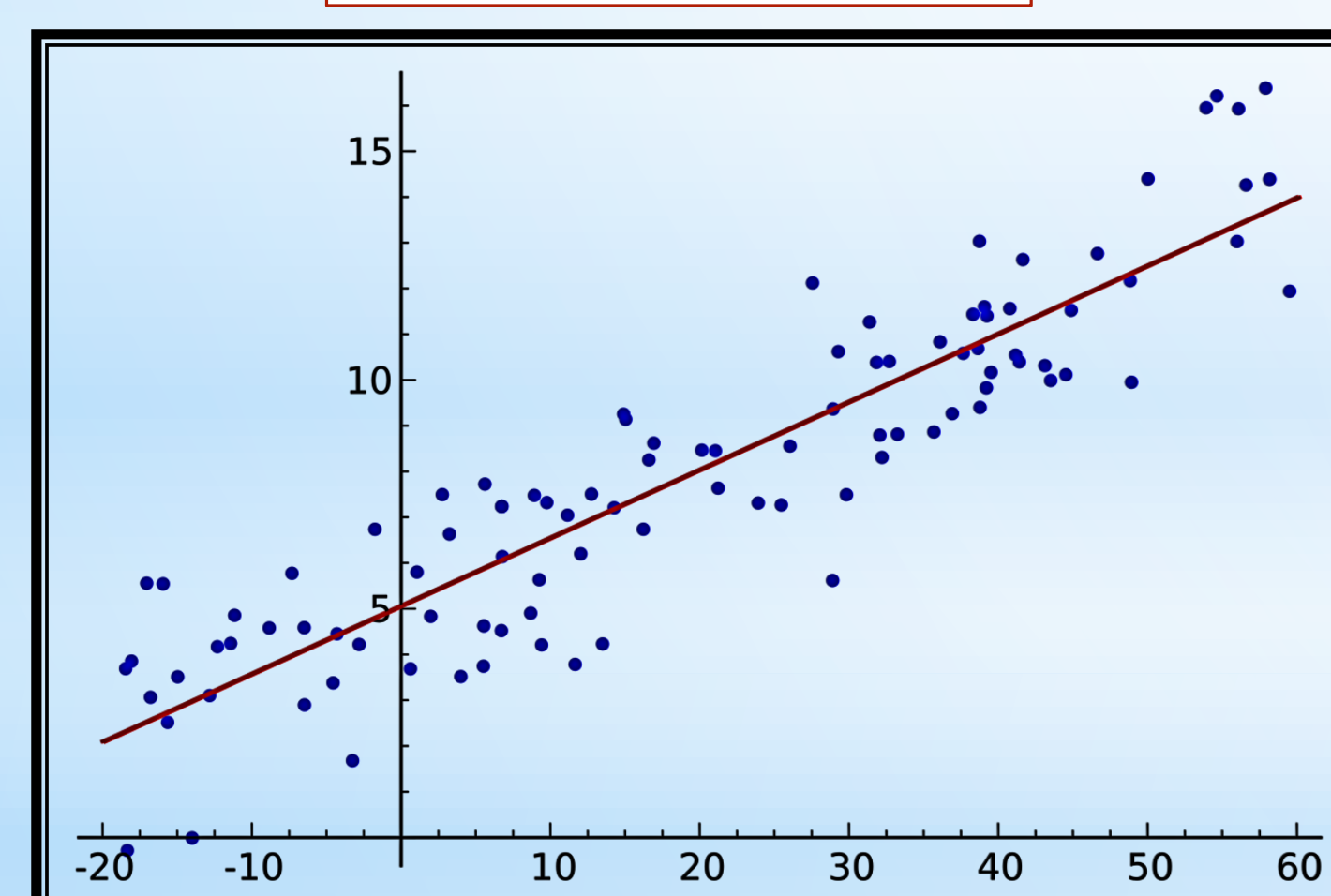
Here, we hypothesize that state-of-the-art deep neural networks can provide further improvement.

Data Structure

- 199,491 patients with 393,260 echocardiograms
- 10 Clinical Variables (age, sex, height ...)
- 90 binary codes (ICD-10¹)
- 50 additional echocardiographic measurements (LV V1 max, LVOT diam...)
- Test with linear regression (LR), random forest (RF), and multiple layer perceptron (MLP)

<Linear Model>

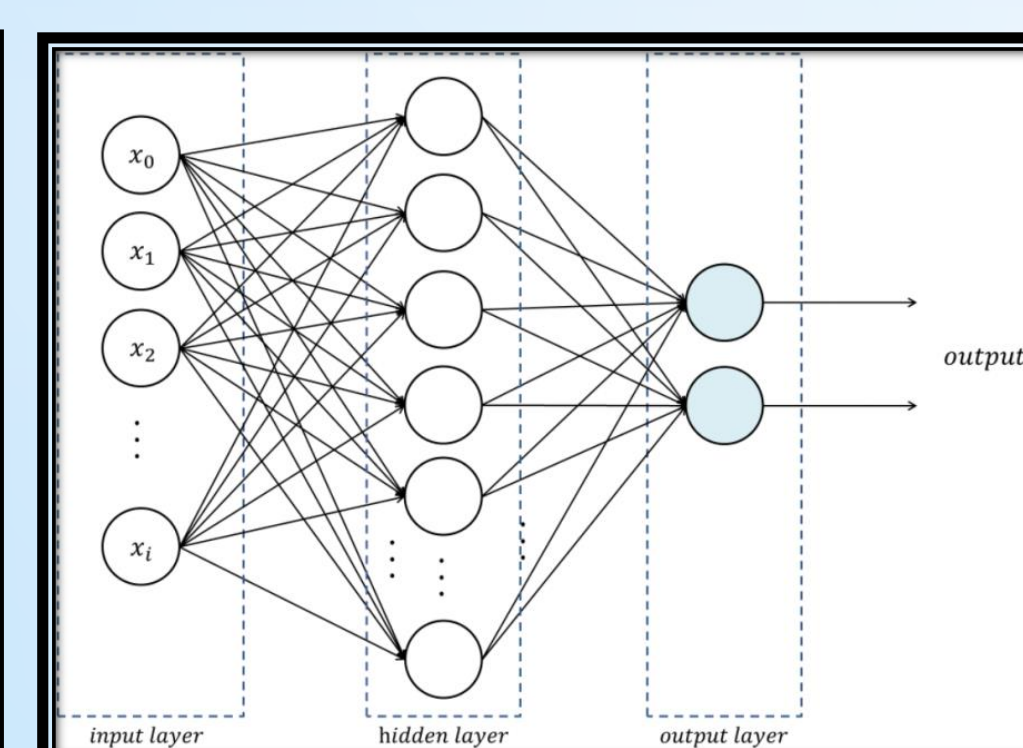
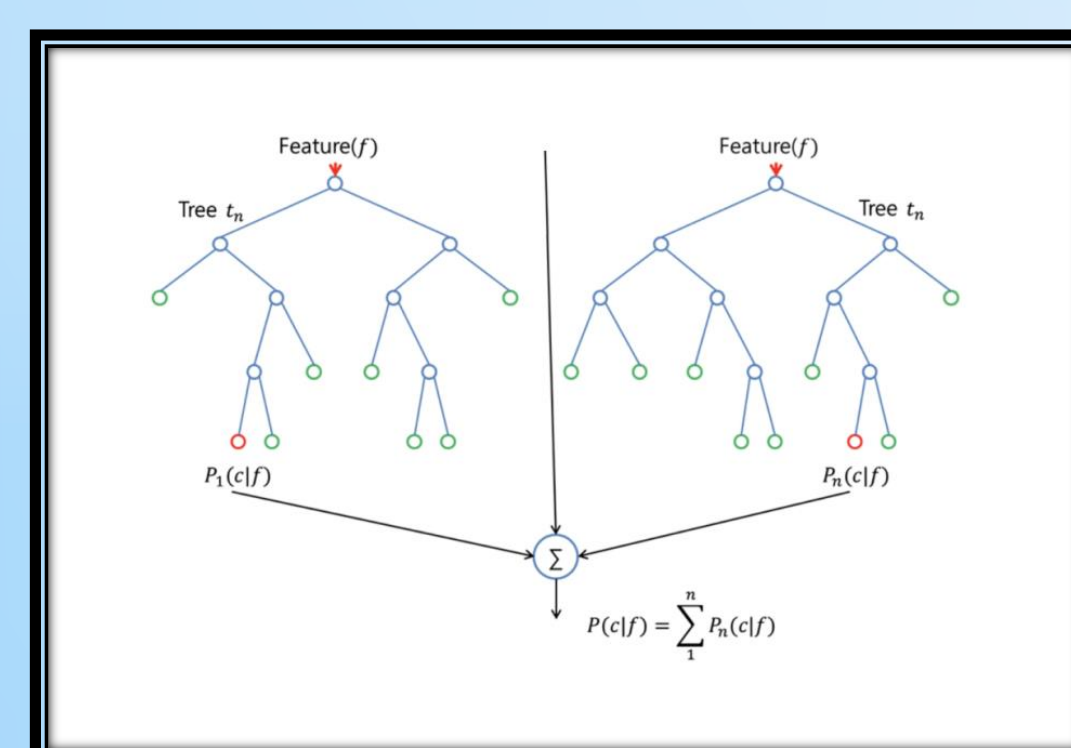
Linear Regression



<Non-Linear Model>

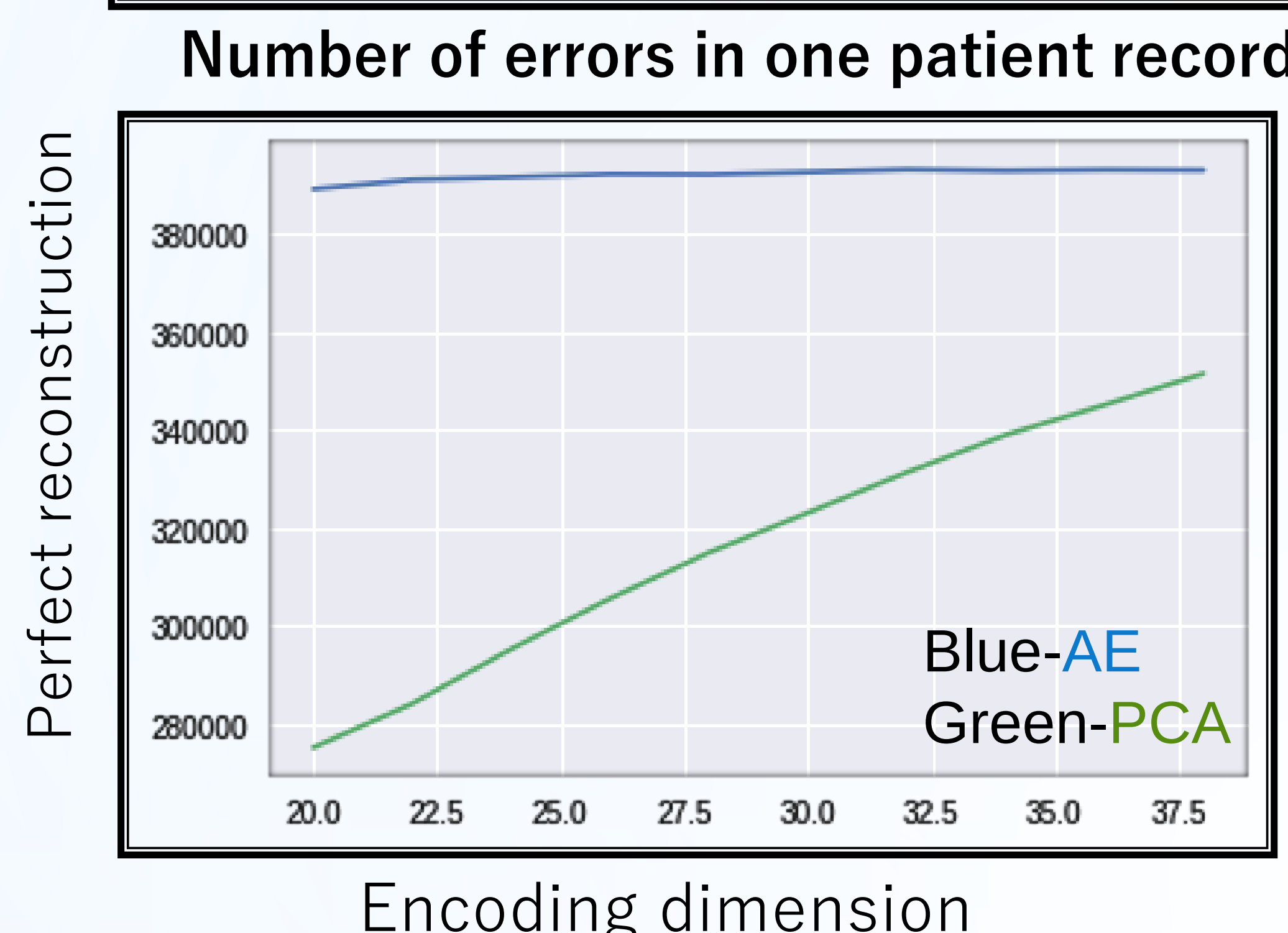
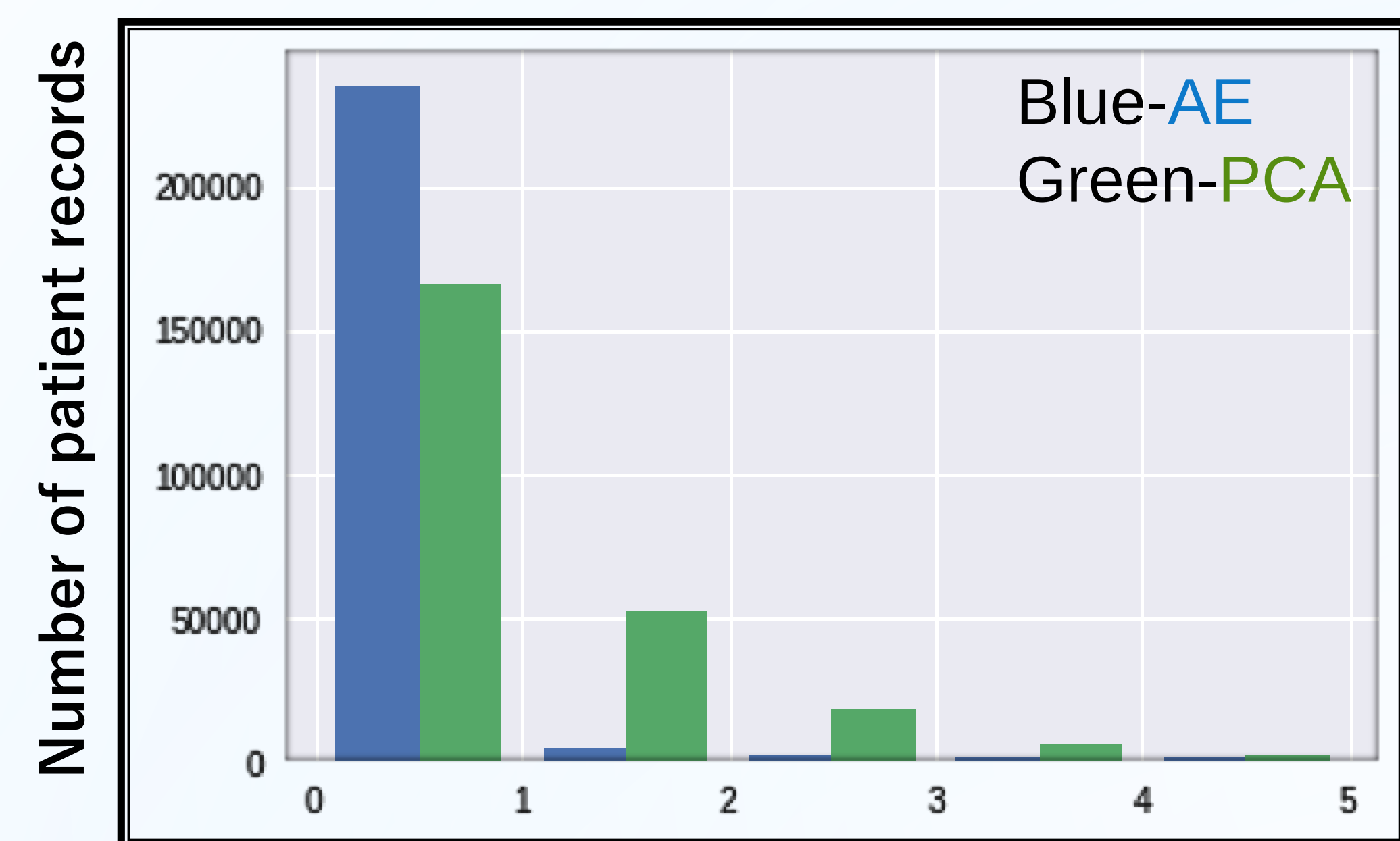
Random Forest

Multiple Layer Perceptron

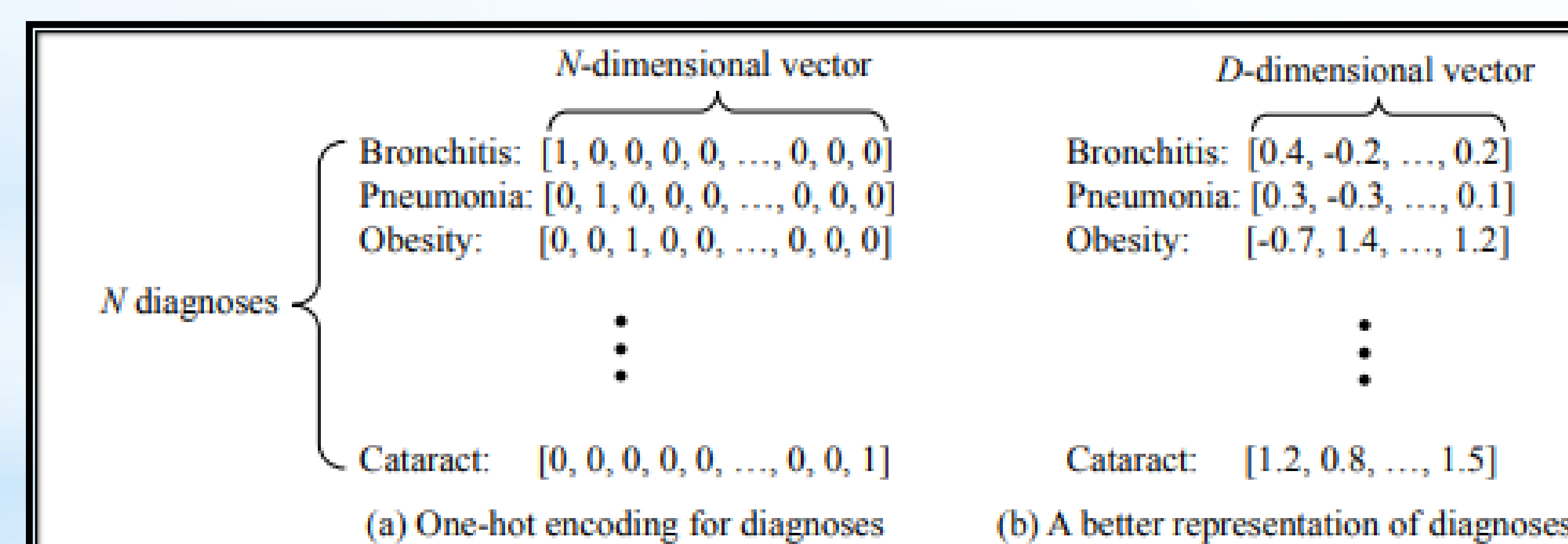


Source: https://en.wikipedia.org/wiki/Linear_regression
<https://towardsdatascience.com/the-random-forest-algorithm-d457d499fcd>
<https://www.cc.gatech.edu/~san37/post/dlhc-fnn/>

Dimension Reduction



Medical Representation Vector

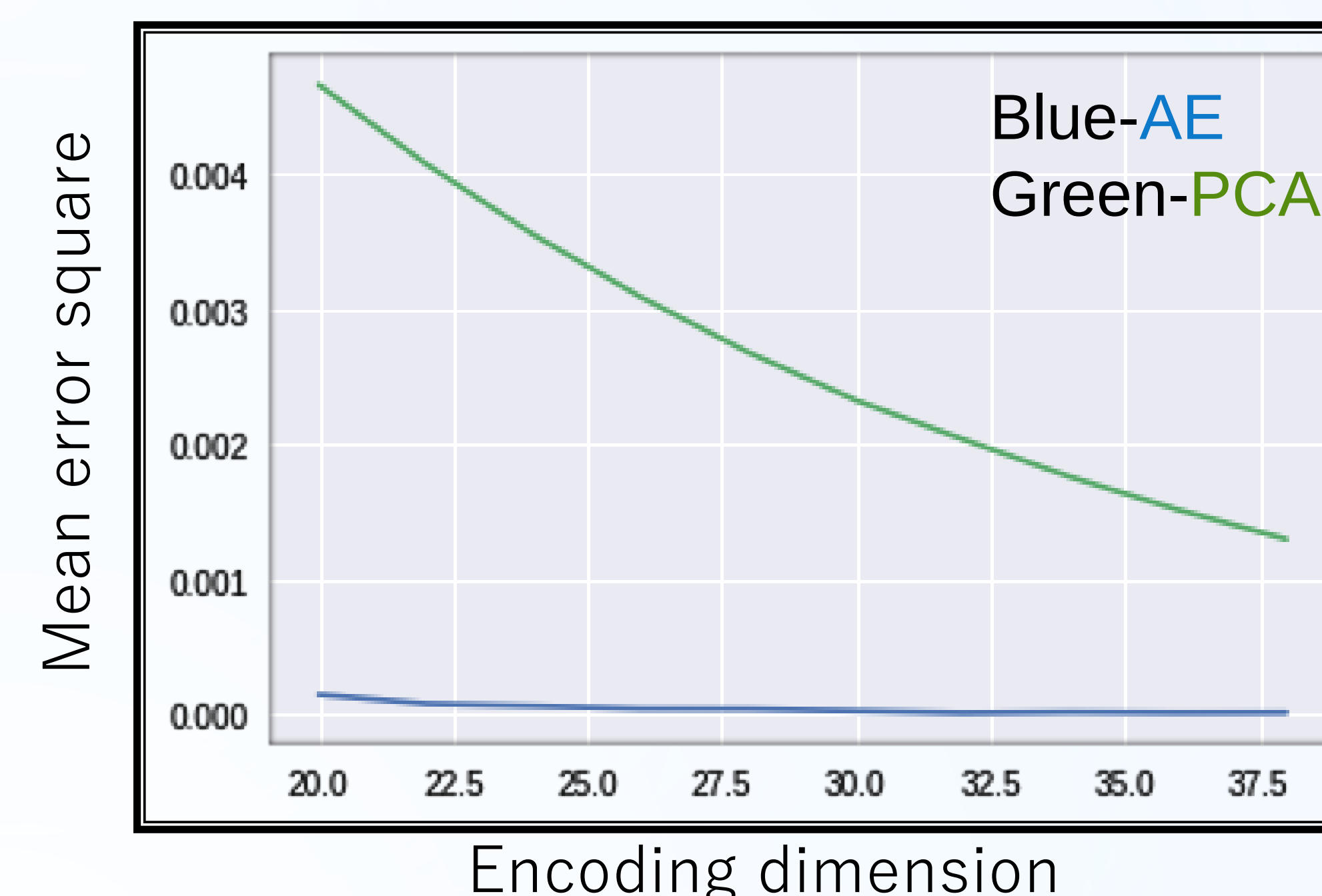


Medical Vector Representation²

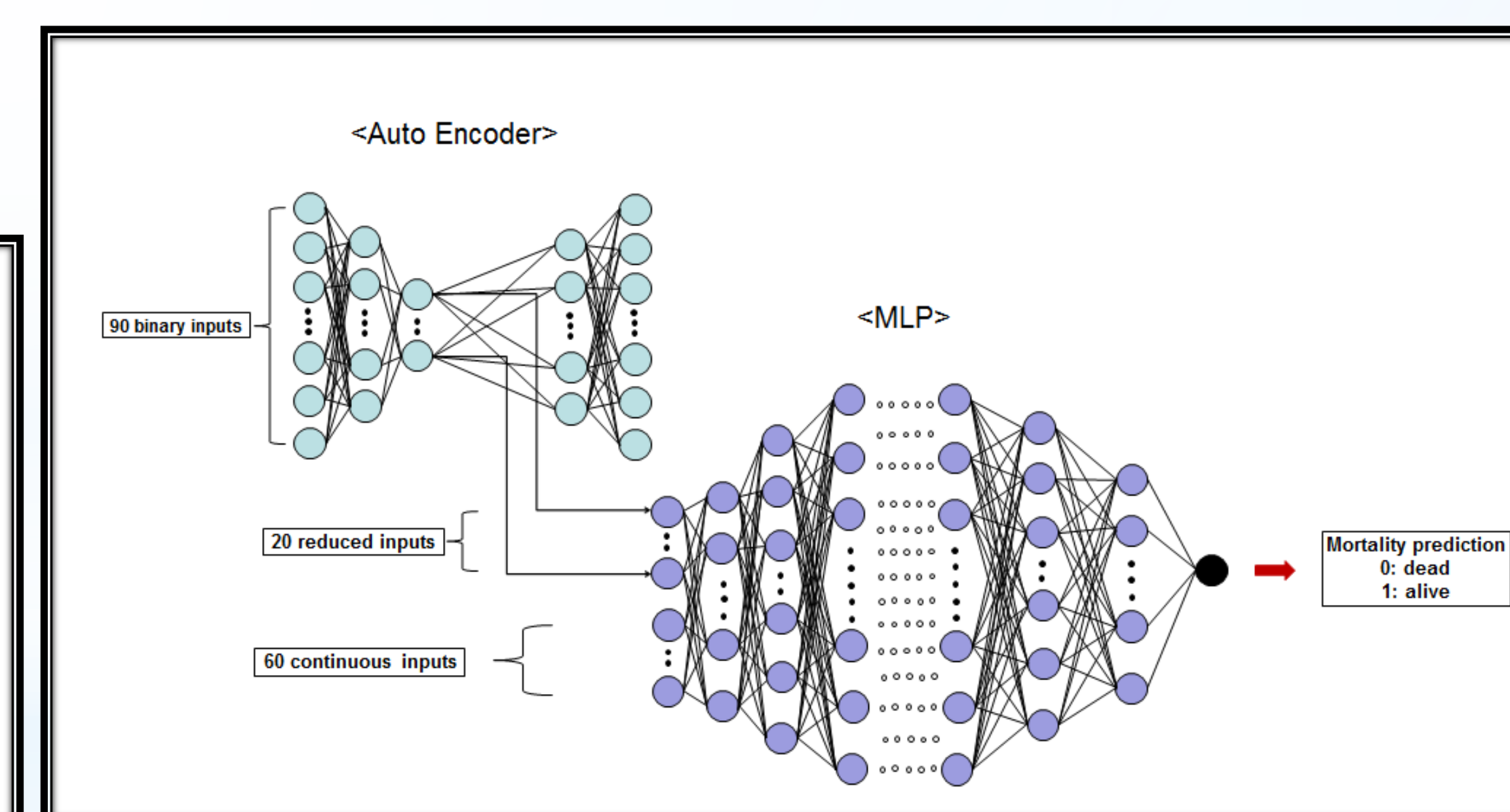
- Feeding 90 dimension (N) binary data to network is not efficient
- Instead, using PCA and AE, we turned 90 dimensions (N) binary data into 20 dimensions (D) continuous data

Source: http://www.sthda.com/english/wiki/wiki.php?id_contents=7851
<https://hackernoon.com/autoencoders-deep-learning-bits-1-11731e200694>

- Principal component analysis (PCA) and neural network auto encoder (AE) are widely used dimensionality reduction techniques.
- PCA is linear, makes assumptions on data distribution. AE provides improved reconstruction quality for non-linear data.
- In this research, we used AE because it has less error and better reconstruction levels than PCA.



Network Pipeline



Results

Original Data Performance

	1 year			5 years		
FRS (8 clinical variables)	0.58			0.61		
Clinical+LVEF+Echo (150 variables)	LR	RF	MLP	LR	RF	MLP
	0.812	0.833	0.822	0.817	0.843	0.837
	[0.81, 0.814]	[0.831, 0.835]	[0.818, 0.825]	[0.815, 0.819]	[0.841, 0.845]	[0.835, 0.839]

Performance after dimension reduction

	5 years		
	LR	RF	MLP
Original	0.817 [0.815, 0.819]	0.843 [0.841, 0.845]	0.837 [0.835, 0.839]
PCA	0.810 [0.807, 0.813]	0.846 [0.844, 0.848]	0.838 [0.834, 0.84]
AE (20)	0.813 [0.810, 0.816]	0.846 [0.844, 0.848]	0.853 [0.851, 0.855]

- RF has the highest AUC for the original data, but after combining AE, MLP reported the highest AUC
- The AUC for LR became lower with PCA or AE
- In conclusion, with more efficient data set, the network has better performance, especially with MLP

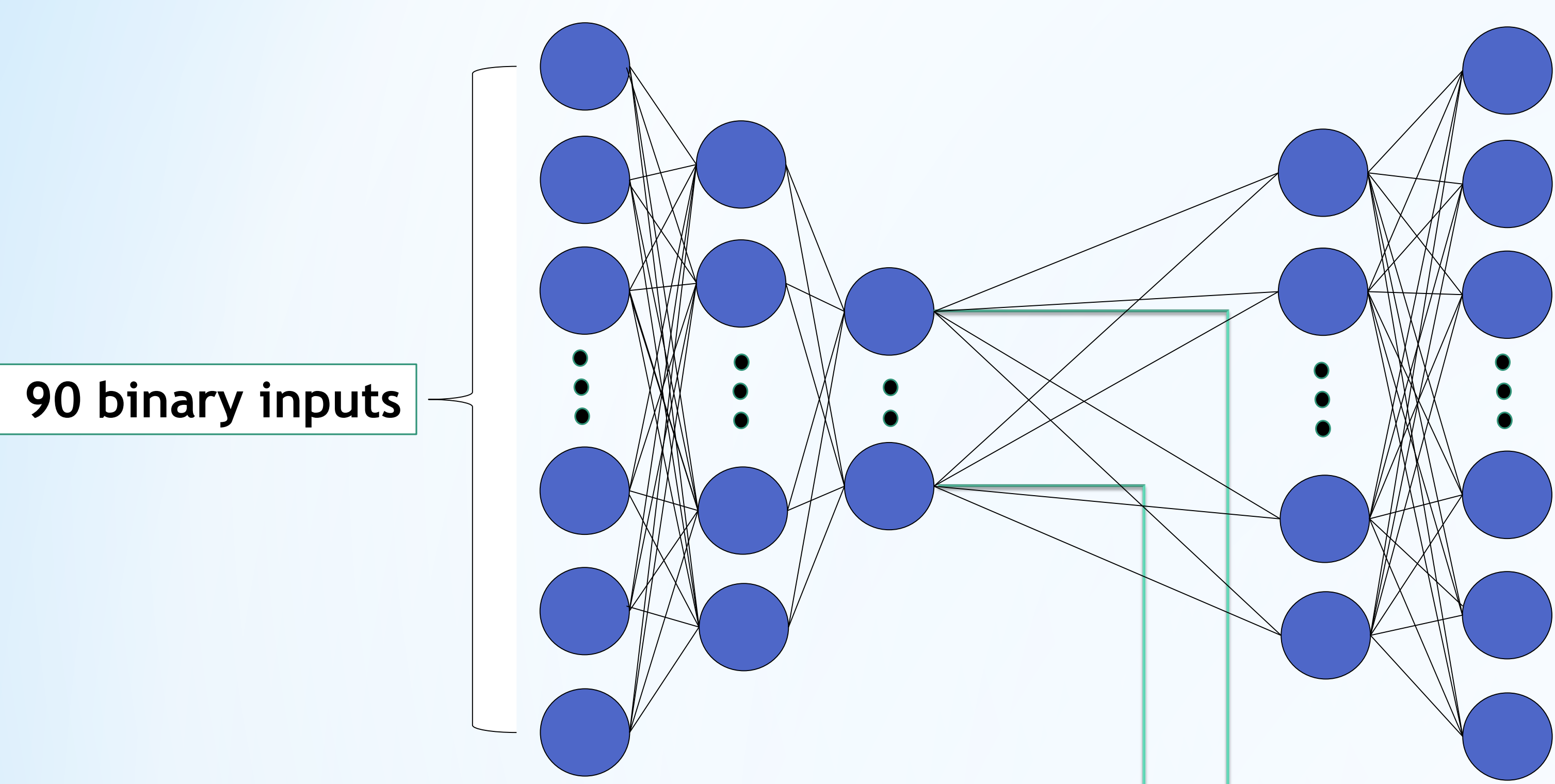
Discussion / Future Plans

- AE with 30 encoding dimensions has the lower error and higher reconstruction rate than 20 encoding dimensions, but we chose 20 because it reported better AUC
- Further improve the performance of MLP, combining with other dimension reduction method

References / Acknowledgements

- International Classification of Diseases version 10
- Choi, E., Schuetz, A., Stewart, W. F., & Sun, J. (02/2016). Medical Concept Representation Learning from Electronic Health Records and its Application on Heart Failure Prediction. *Eprint ArXiv:1602.03686*.
- Samad et al. 2018
- Geisinger Summer Undergraduate Research Program (SURP)
- Cardiac Imaging Technology Lab

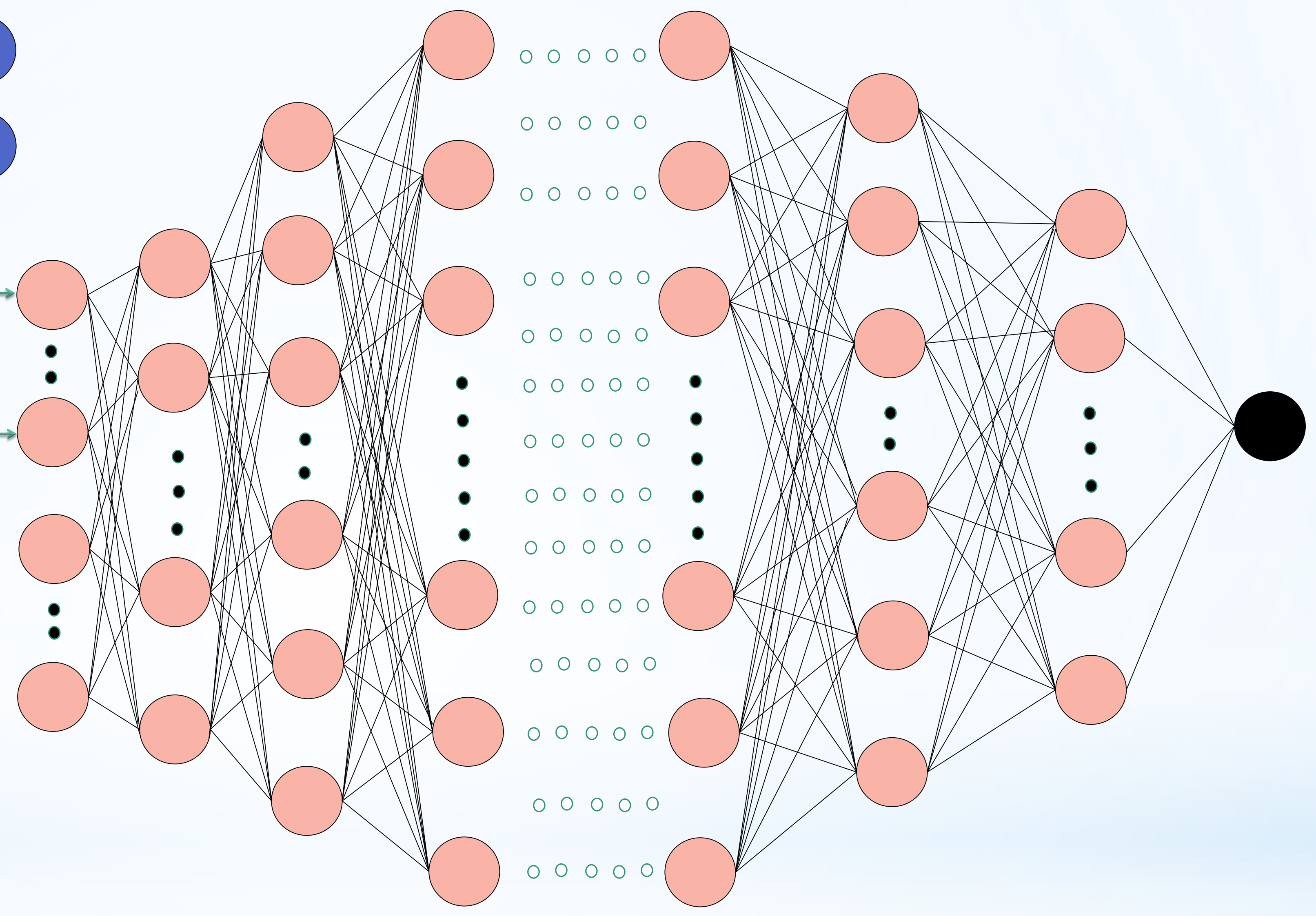
<Auto Encoder>



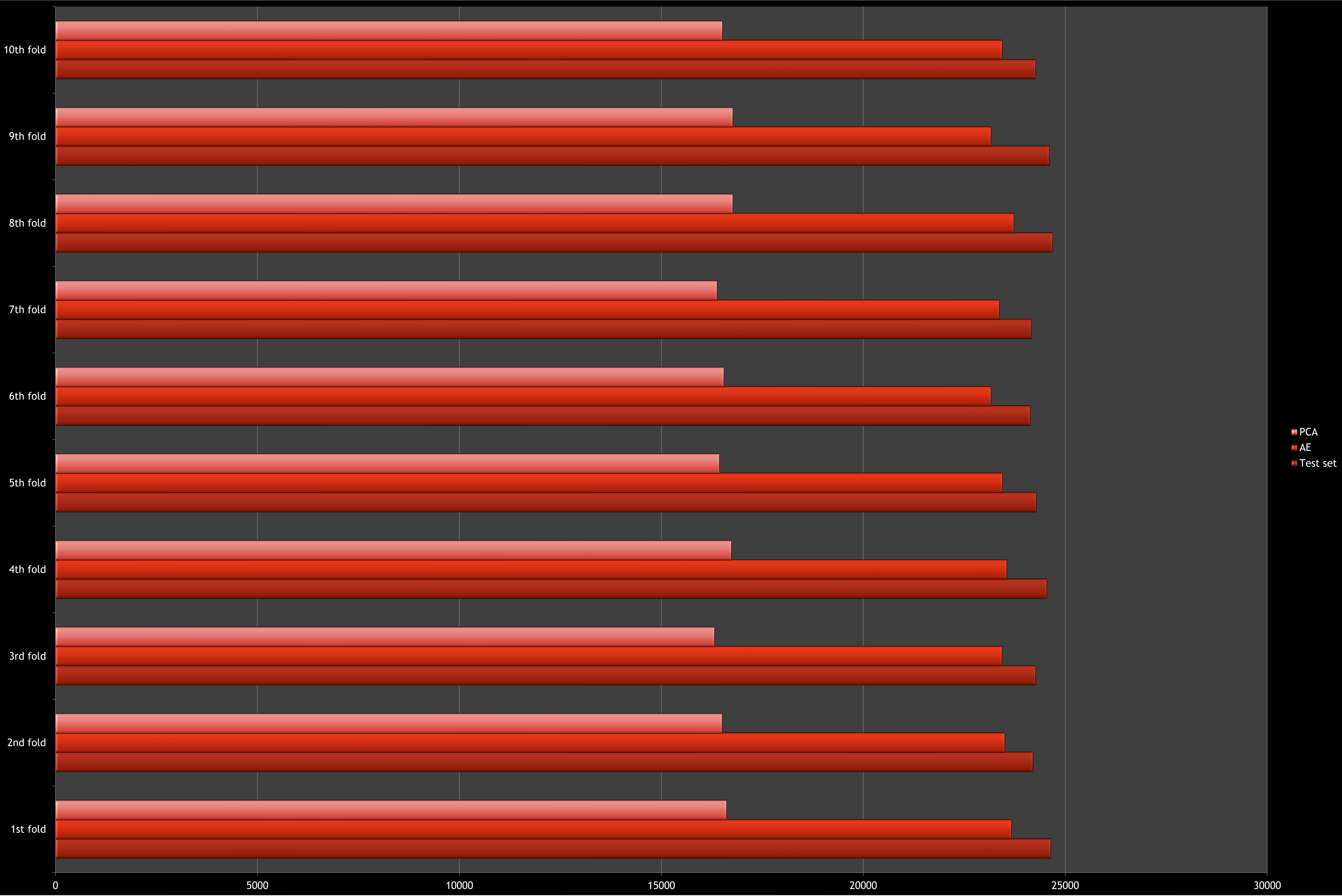
20 reduced inputs

60 continuous inputs

<MLP>



Mortality prediction
0: dead
1: alive



	1 year			5 years		
FRS (8 clinical variables)	0.58			0.61		
	LR	RF	MLP	LR	RF	MLP
Clinical+ LVEF+ Echo	0.812	0.833	0.822	0.817	0.843	0.837
	[0.81, 0.814]	[0.831, 0.835]	[0.818, 0.825]	[0.815, 0.819]	[0.841, 0.845]	[0.835, 0.839]

	5 years		
	LR	RF	MLP
Original	0.816 [0.813,0.819]	0.843 [0.842, 0.846]	0.837 [0.835, 0.839]
PCA	0.810 [0.807, 0.813]	0.846 [0.844, 0.848]	0.838 [0.834, 0.84]
AE (20)	0.813 [0.810, 0.816]	0.846 [0.844, 0.848]	0.853 [0.851, 0.855]

Original Data

PCA reconstruction

AE reconstruction

