Mortality Prediction on MIMIC-III Clinical Datasets Using RNN (GRU) and GNN

Poornima Haridas (ph1391)
Yuan Ding (yd1400)

05/04/2020

NEW YORK UNIVERSITY

Motivation and Related Work

Predicting the mortality using the data from 24hr / 48hr windows after ICU admission:

- Help with the assessment of severity of illness
- Guide decision making in terms of drastic measures required to save a patient

Why deep learning:

- DL models **outperform** ML models in ICU scoring systems especially when a large number of **raw** clinical time series data is used as input features

Related work:

1. Purushotham, S., Meng, C., Che, Z., & Liu, Y. (2017). Benchmark of deep learning models on large healthcare mimic datasets. arXiv preprint arXiv:1710.08531.
2. Zhu, W., & Razavian, N. (2019). Graph Neural Network on Electronic Health Records for Predicting Alzheimer's Disease. arXiv preprint arXiv:1912.03761.

Data (1): Preprocessing

Medical Information Mart for Intensive Care (MIMIC) III (Johnson et al., 2016):

- De-identified, comprehensive clinical data of patients admitted to an ICU at the Beth Israel Deaconess Medical Center (BIDMC) in Boston during 2001 to 2012
- We built a relational database “mimic” with 26 raw tables using **PostgreSQL**

Data Preprocessing:

- **Cohort selection** → first admission of patients >15 years old
- **Inconsistent units** (e.g. “dose”, “mg”) and data type (string/numeric) → unify
- **Multiple recordings** at the same time → take the mean
- **Value recorded as a range** at a certain time → take the median
- **Missing values** → forward-backward imputation; mean imputation

Item	Mortality ratio
In-hospital	0.112
2-day	0.018
3-day	0.014

Item	Overall	Dead at hospital	Alive at hospital
# admissions	**36,093**	4029	32,064
Age	65.86 [15.06, 89.09]	73.85 [15.18, 89.05]	64.98 [15.06, 89.09]
Gender (Female)	15607 (43.24%)	1866 (46.31%)	13749 (42.88%)
Origin: Medical	29,867	3330	26,537
Origin: Emergency	6226	699	5527

Data (2): Feature Extraction

Feature Set A (“processed”):
12 sequential + 5 static features
- Drop outliers according to medical knowledge
- Merge relevant features

Feature Set B (“raw”):
15 sequential + 5 static features
- Do not remove outliers, only drop values below 0
- Use separate raw features (e.g. consider PaO2 and FiO2 as individual features instead of calculating the PF-ratio)

Generate time series:
- Extracted the data from first 24 / 48 hours of ICU stay
- Each feature is sampled per 1 hour

	24 hrs	48 hrs
Feature Set A	36093*(12*24+5)	36093*(12*48+5)
Feature Set B	32290*(12*24+5)	32290*(12*48+5)

Table: 17 features used in Feature Set A

sequential

Item	Table
Glasgow coma scale	chartevents
Systolic blood pressure	chartevents
Heart rate	chartevents
Body temperature	chartevents
Pao2 / fio2 ratio	chartevents
Urine output	outputevents
Serum urea nitrogen level	labevents
White blood cells count	labevents
Serum bicarbonate level	labevents
Sodium level	labevents
Potassium level	labevents
Bilirubin level	labevents
Age	icustays
Acquired immunodeficiency syndrome (AIDS)	diagnoses_icd
Hematologic malignancy	diagnoses_icd
Metastatic cancer	diagnoses_icd
Admission type	admissions

Model(1): GRU

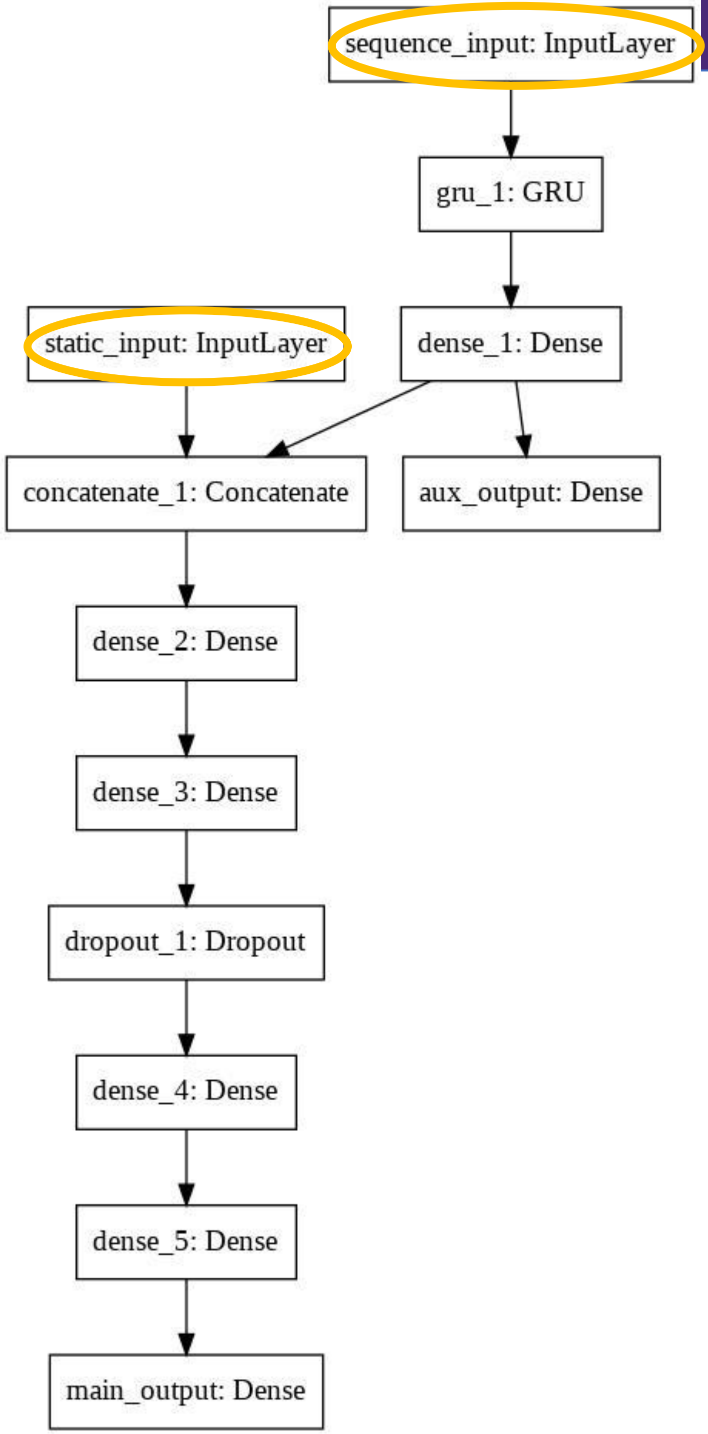
- Successful at modeling sequential data
- Use Keras package

Training:

- **Upsampling** to deal with unbalanced dataset
- **Regularization:** dropout layer, l_2 regularizer
- GRU output dimension: 24 for 2/3-day mortality, 128 for in-hospital mortality
- **Activation** function: sigmoid
- **Loss** function: binary cross entropy
- **Optimizer**: Adam

Parameters tuned:

- Learning rate: [0.001, 0.0001, 0.00001]
- Dropout rate: [0.1, 0.2, 0.3, 0.5]
- Parameter for l_2 regularizer: [0.01,0.02]



Model(2): GNN

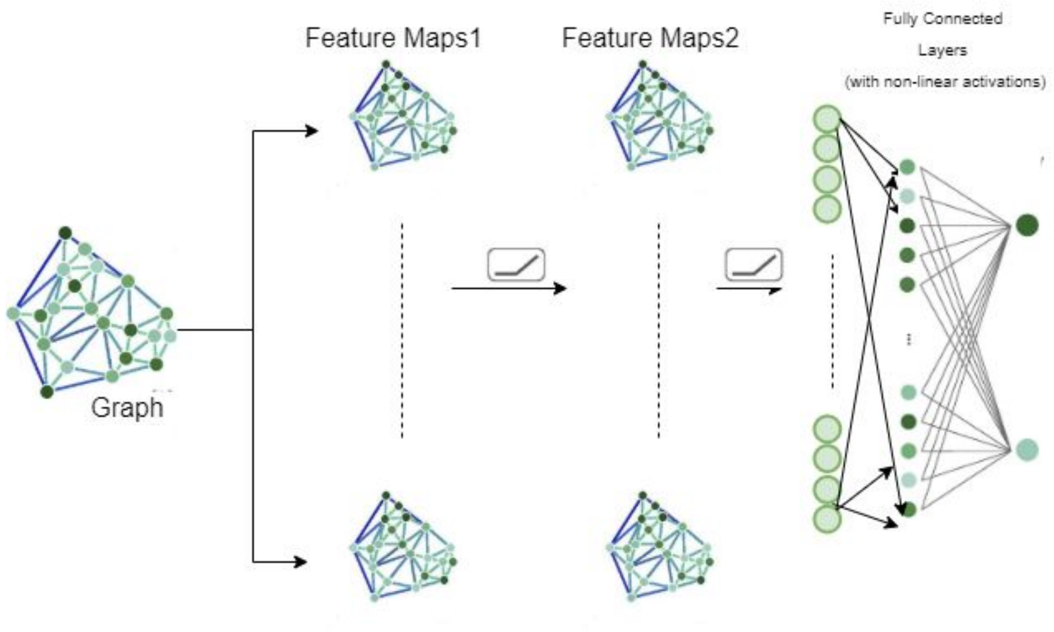
- GNN is a network that takes as input a **graph structure** and utilizes the structural information of the graph for learning.
- Used torch_geometric.nn package
- We used Graph Convolutional layers

Training:

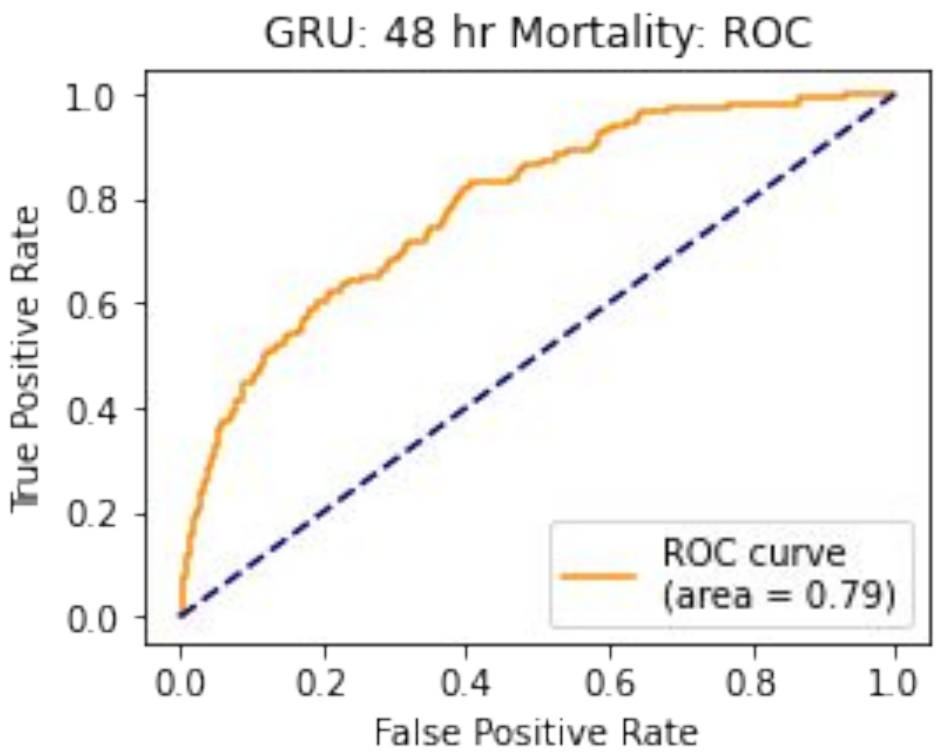
- Five layers:

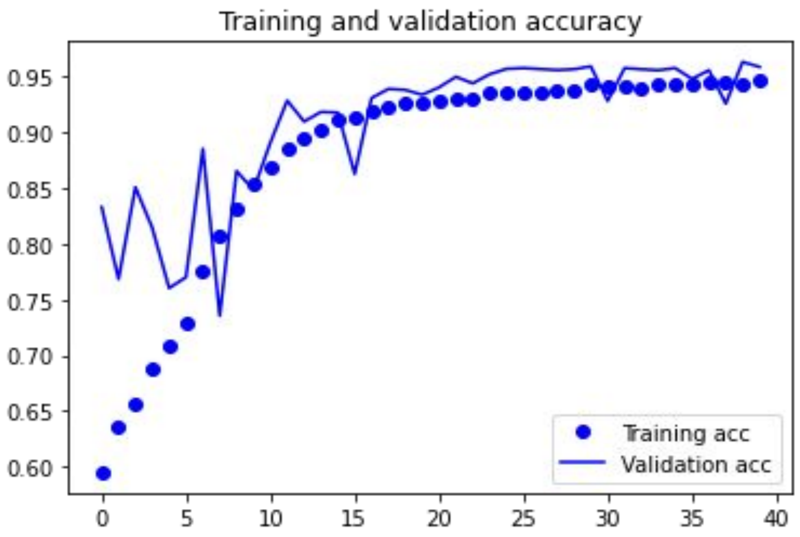
Graph Convolutional layer: 293→256
Graph Convolutional layer: 256→256
FC 1: 256→128
FC 2: 128→64
FC 3: 64→2

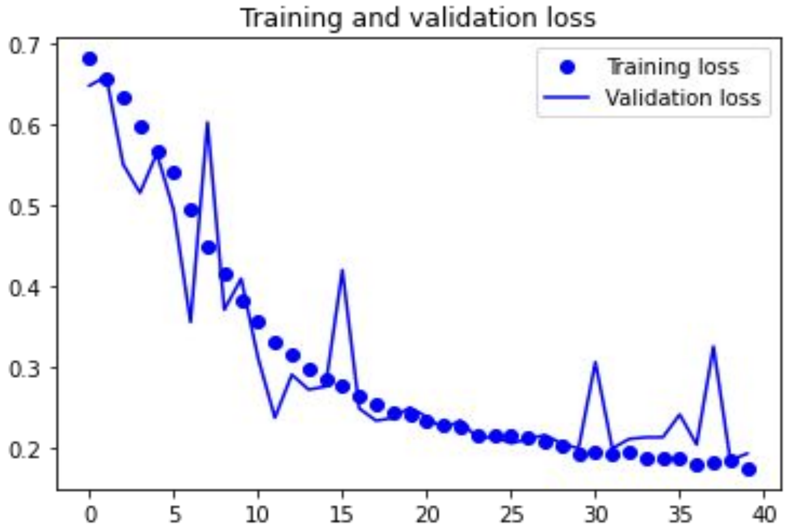
- Loss function: binary cross entropy
- Optimizer: Adam



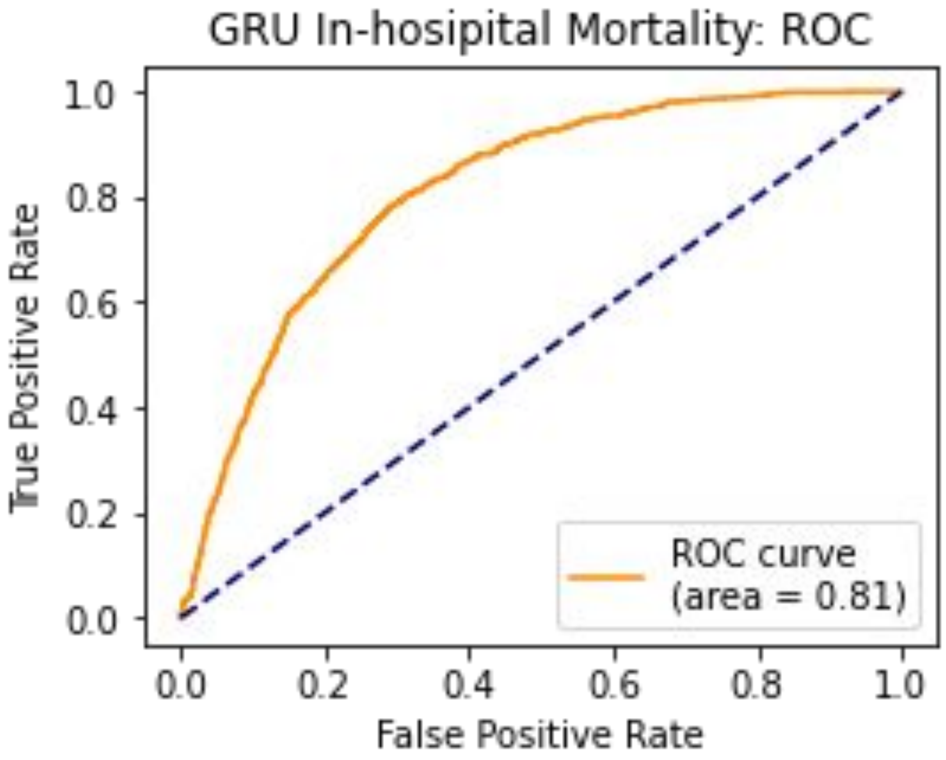
Results: GRU(set A, 24hrs) (1)

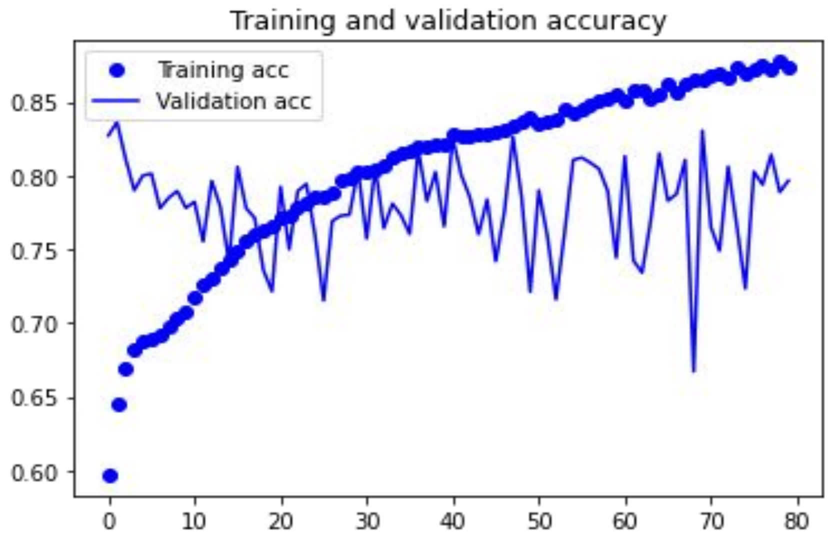


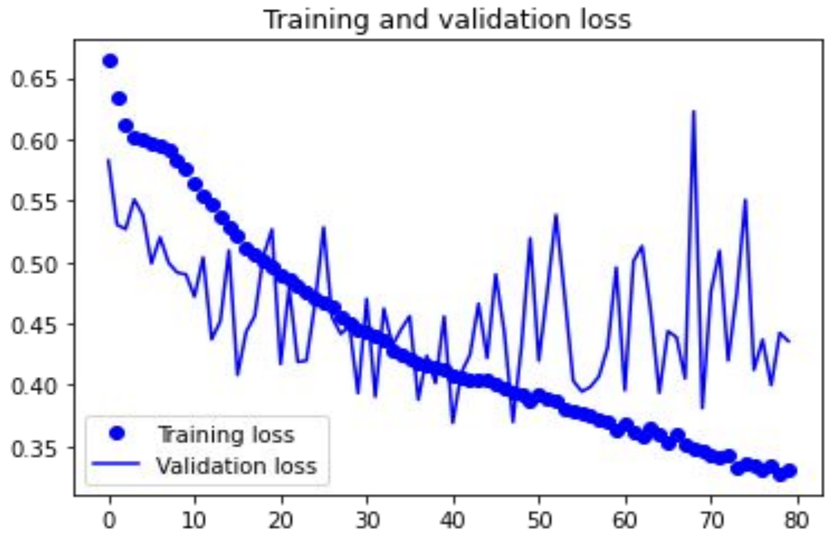




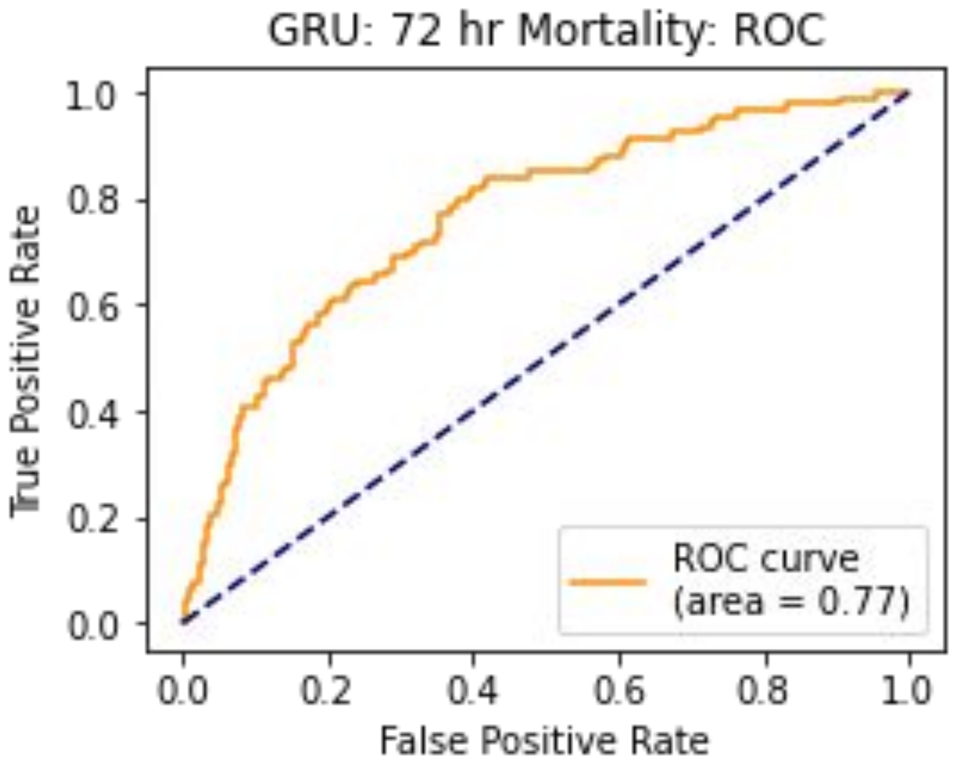
Results: GRU(set A, 24hrs) (2)

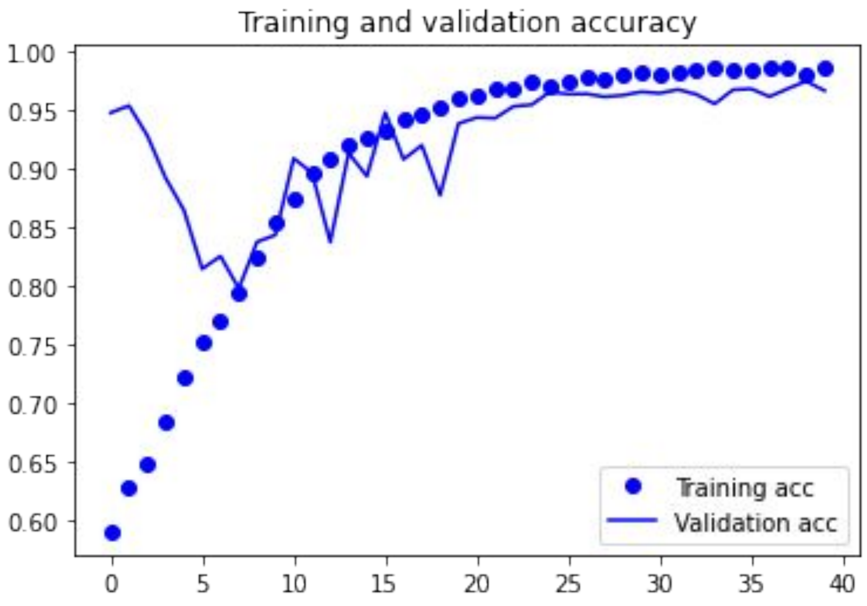


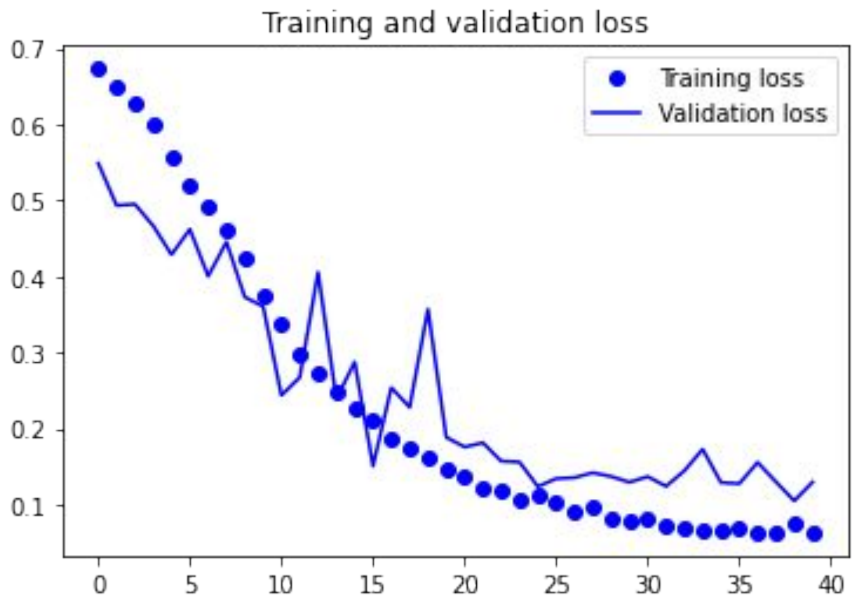




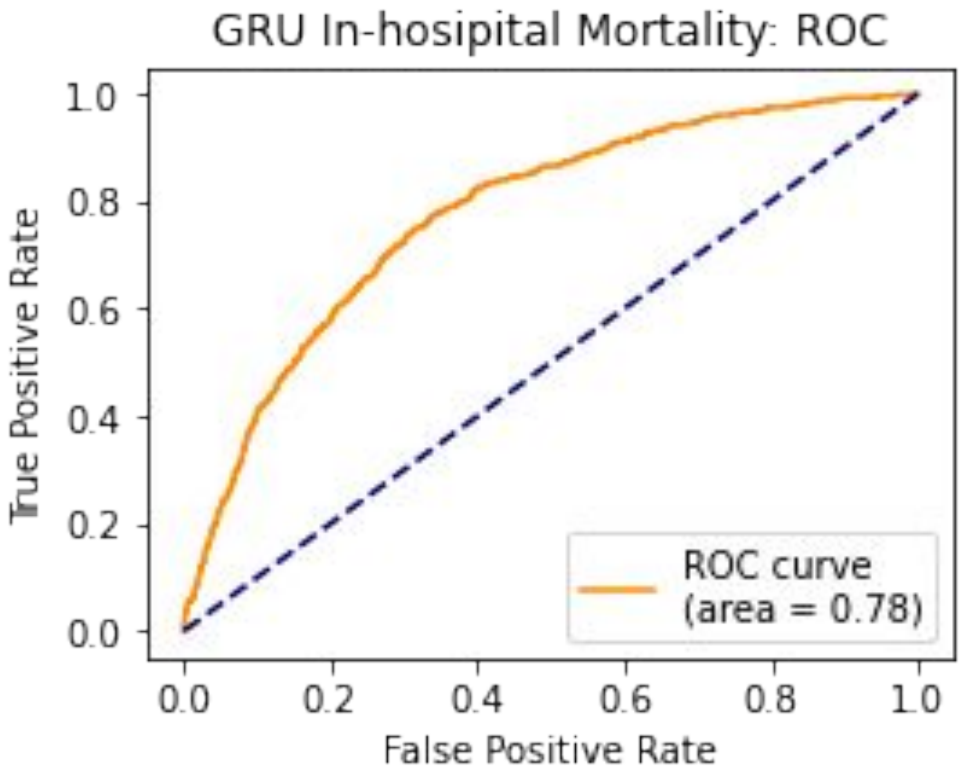
Results: GRU(set A, 48hrs) (1)

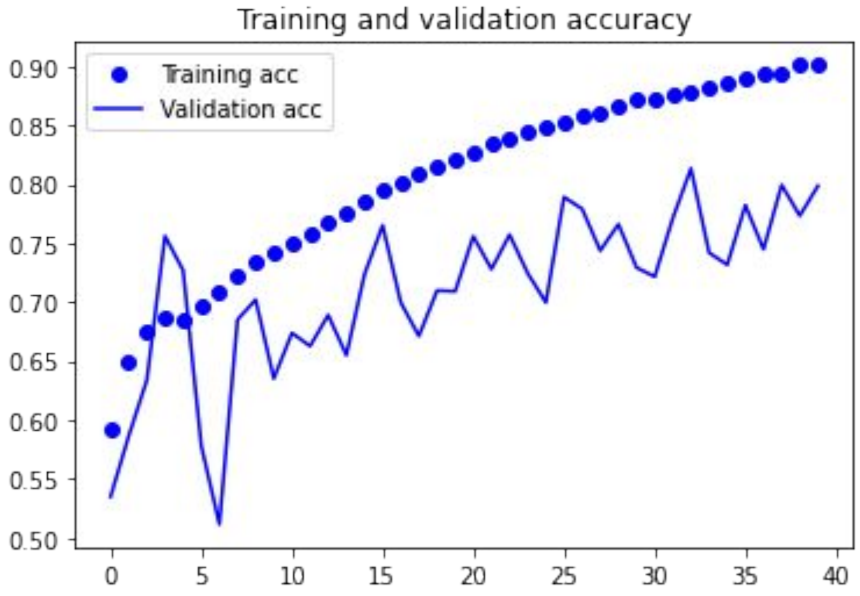


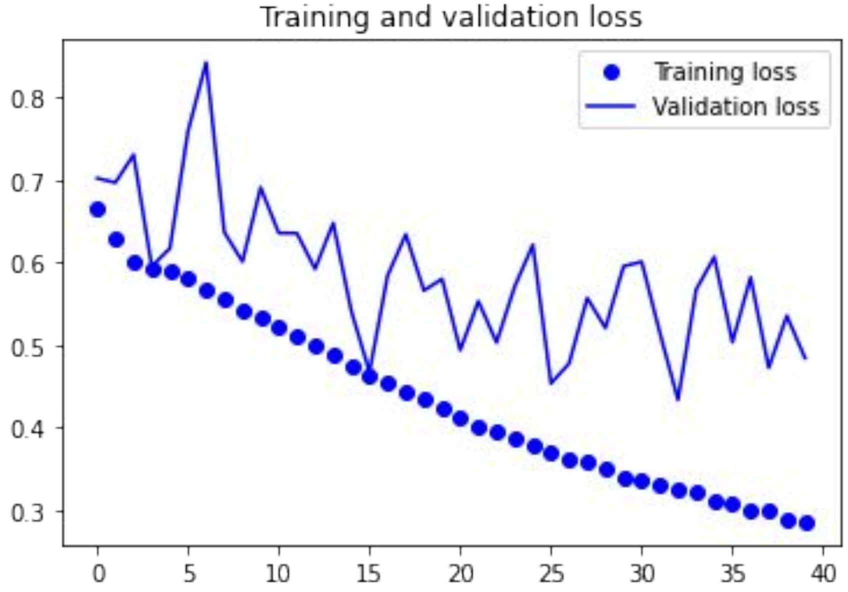




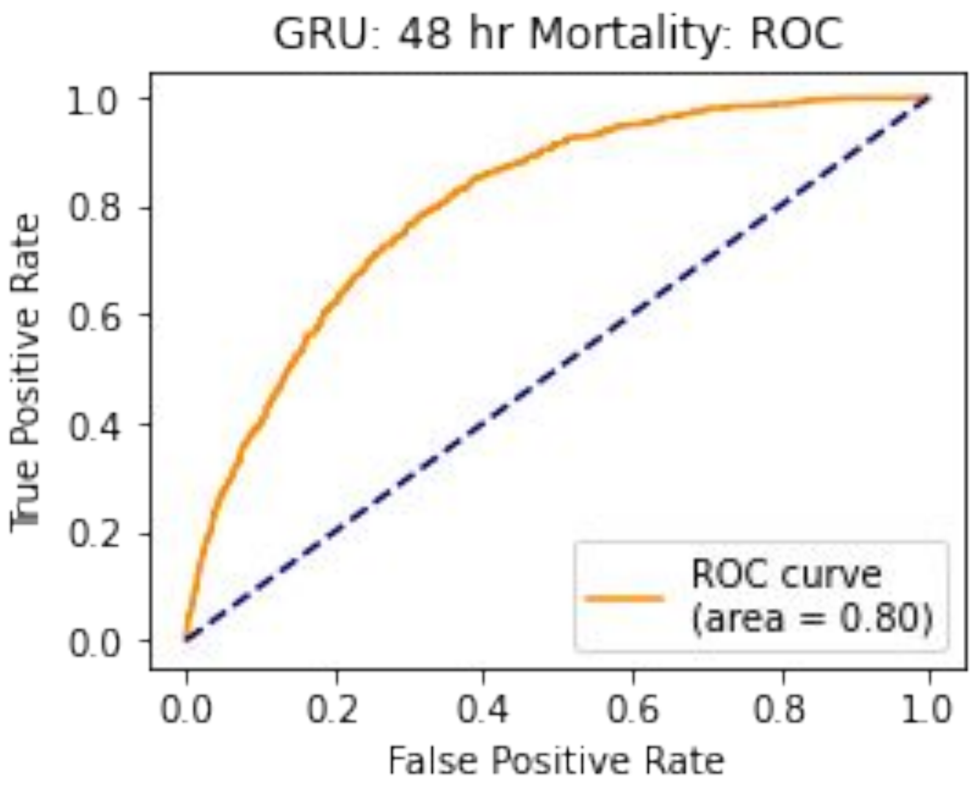
Results: GRU(set A, 48hrs) (2)

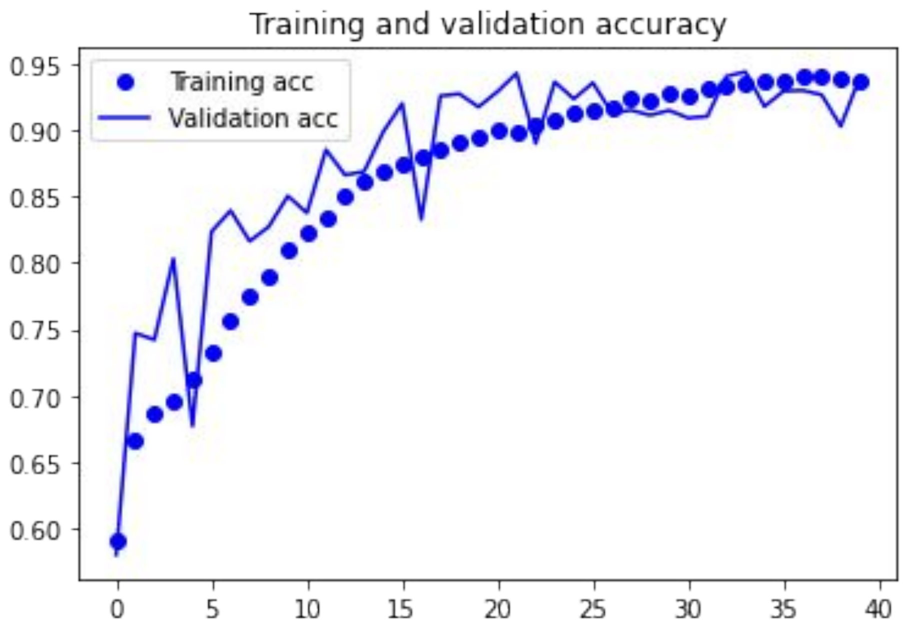


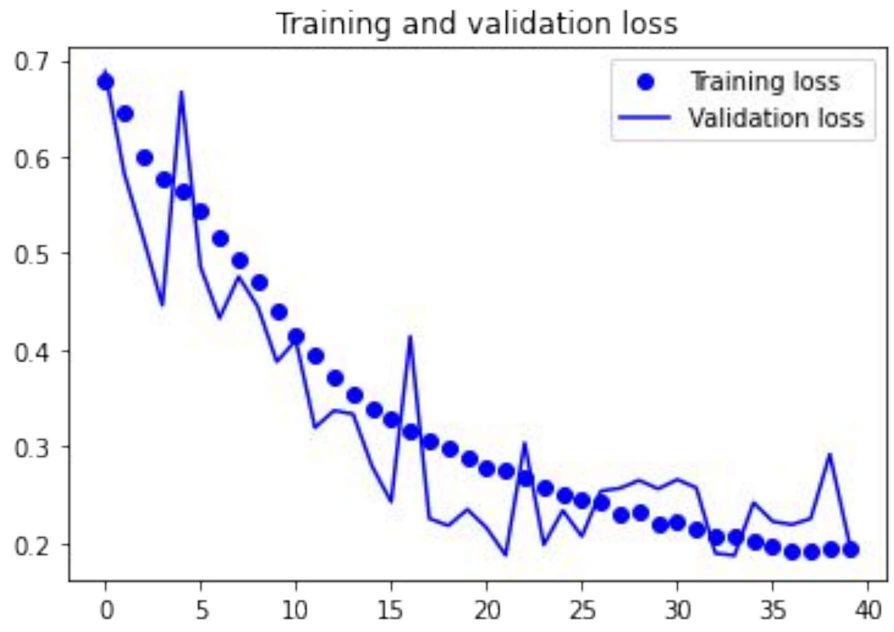




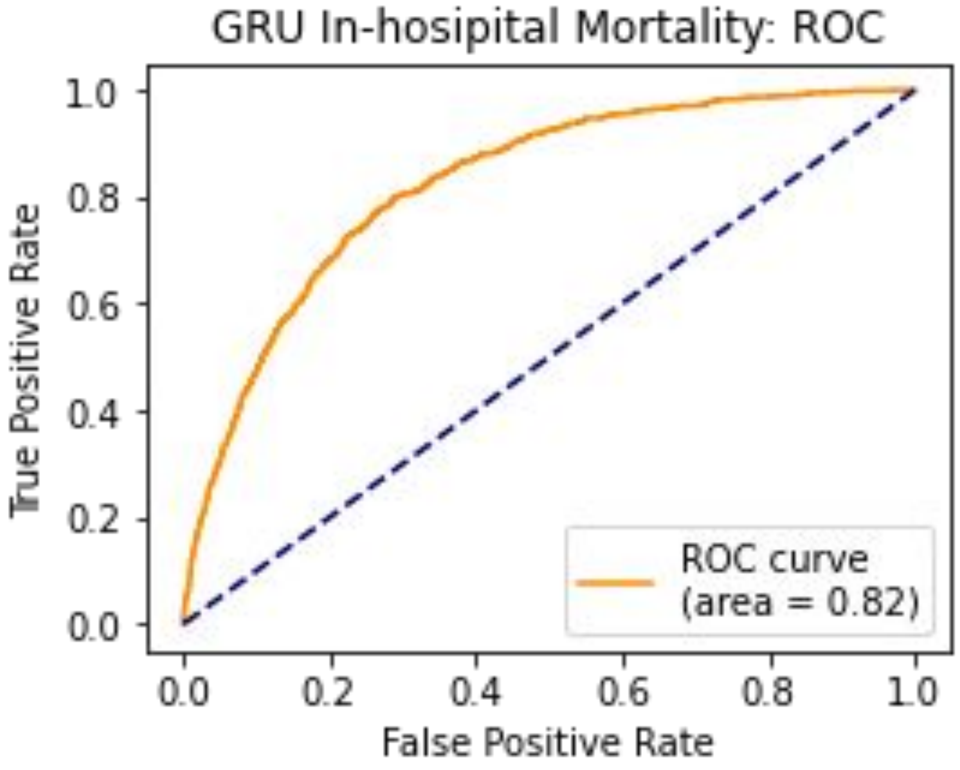
Results: GRU(set B, 24hrs) (1)

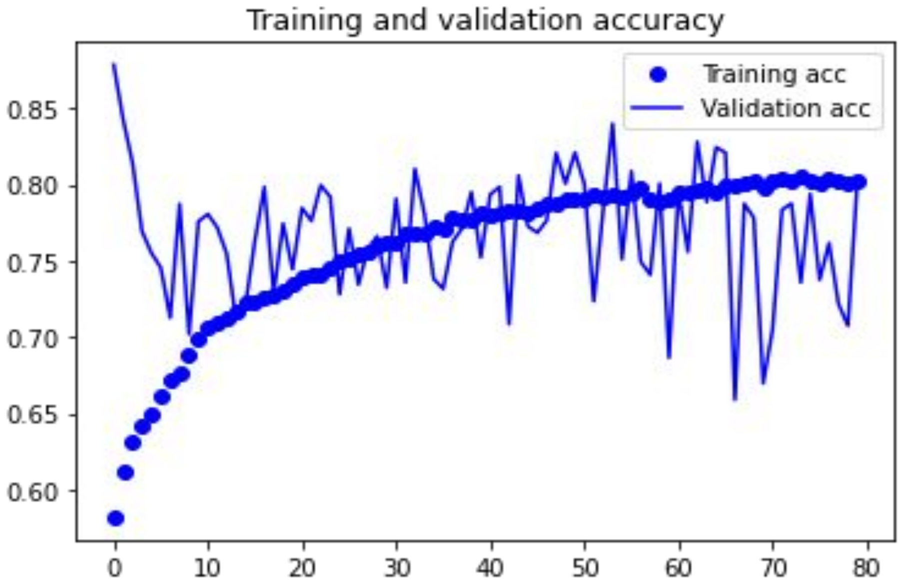


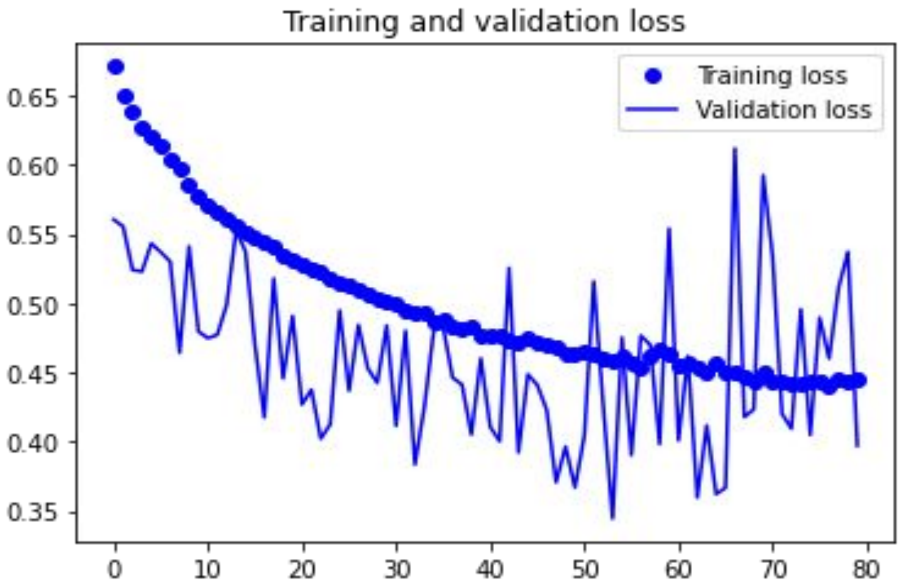




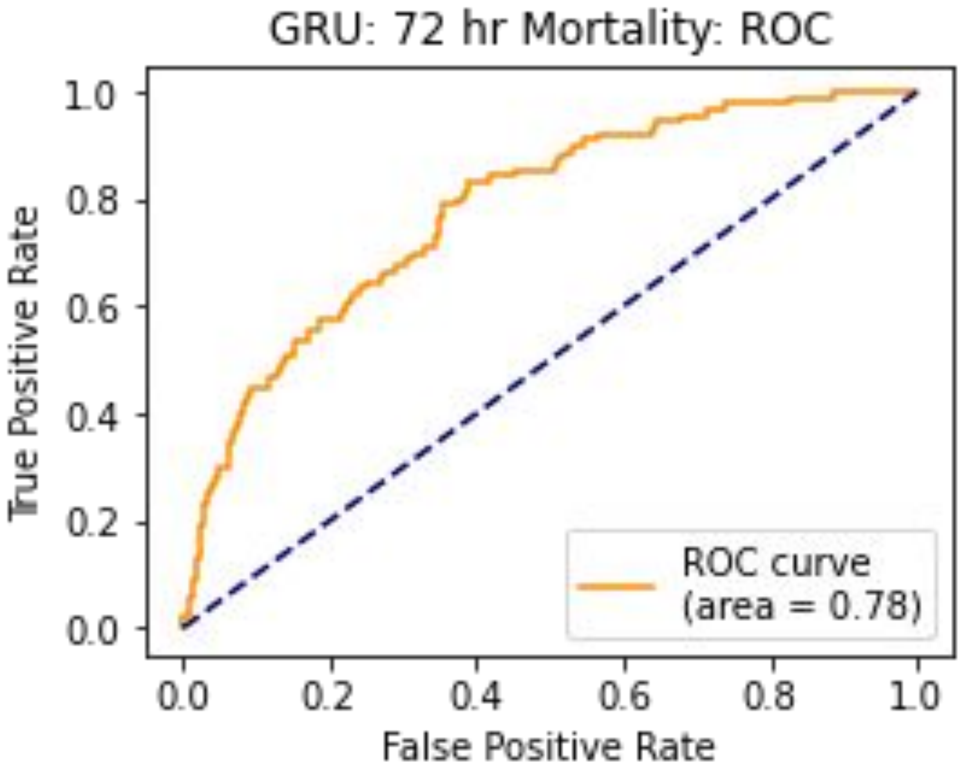
Results: GRU(set B, 24hrs) (2)

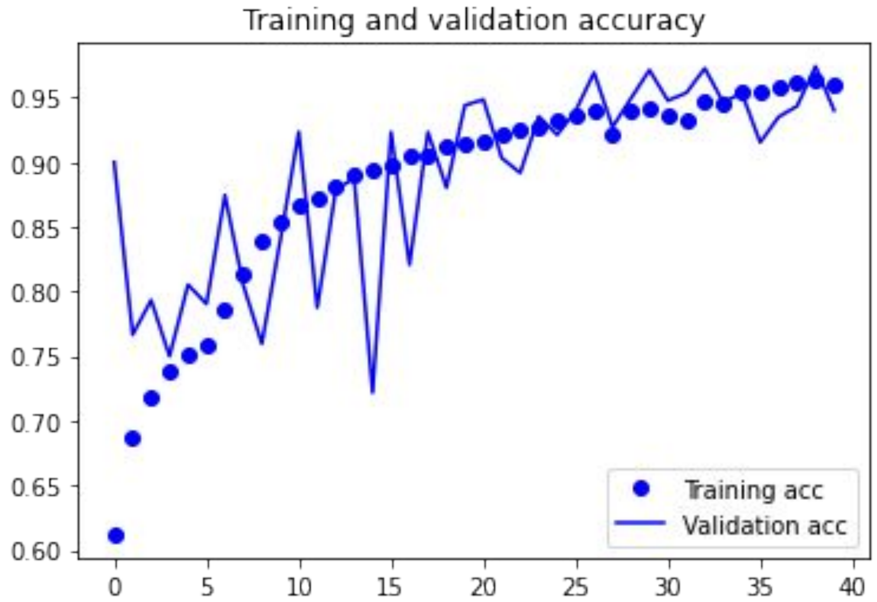


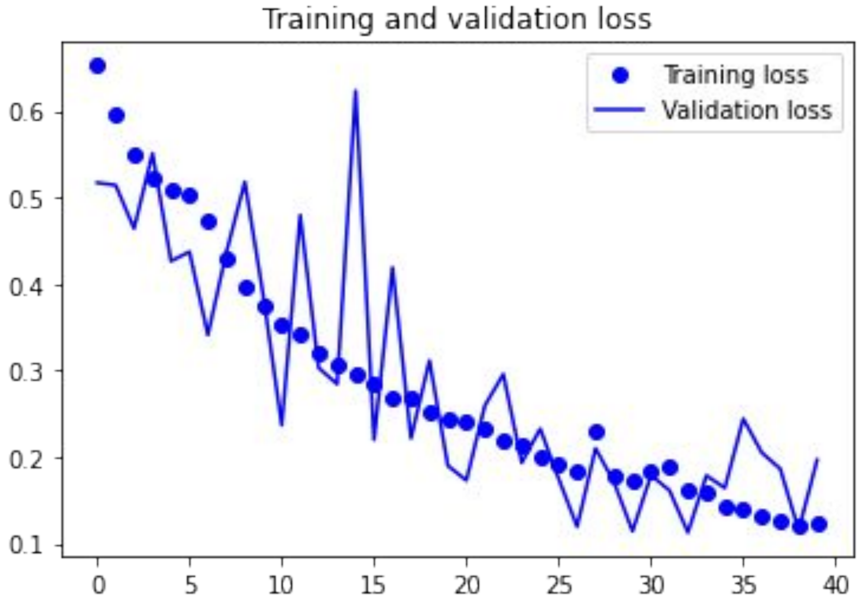




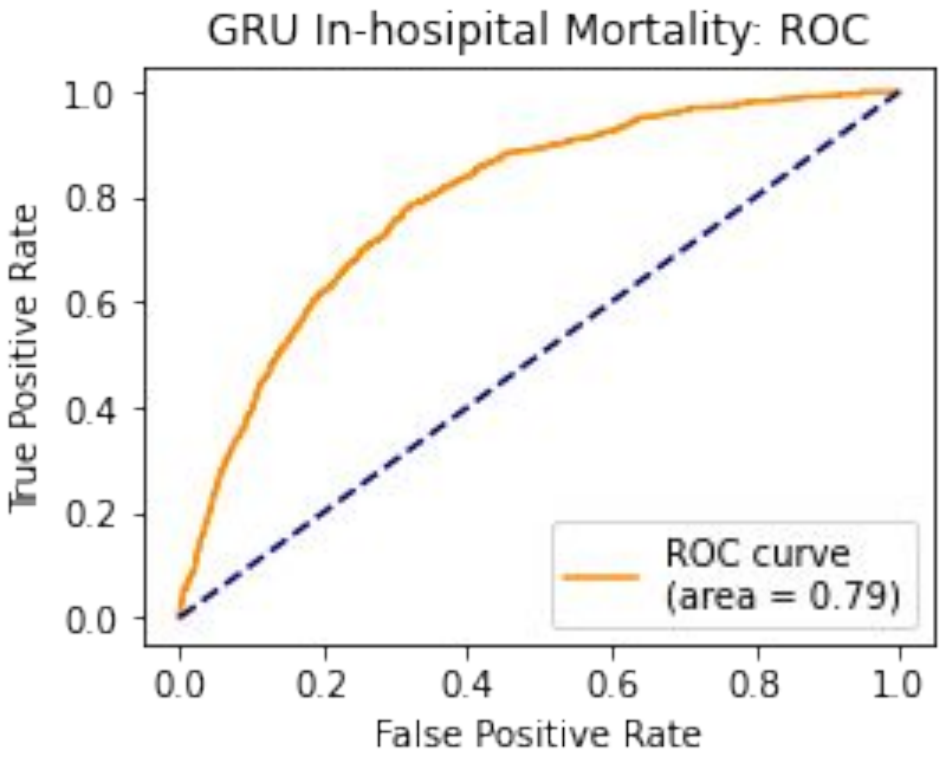
Results: GRU(set B, 48hrs) (1)

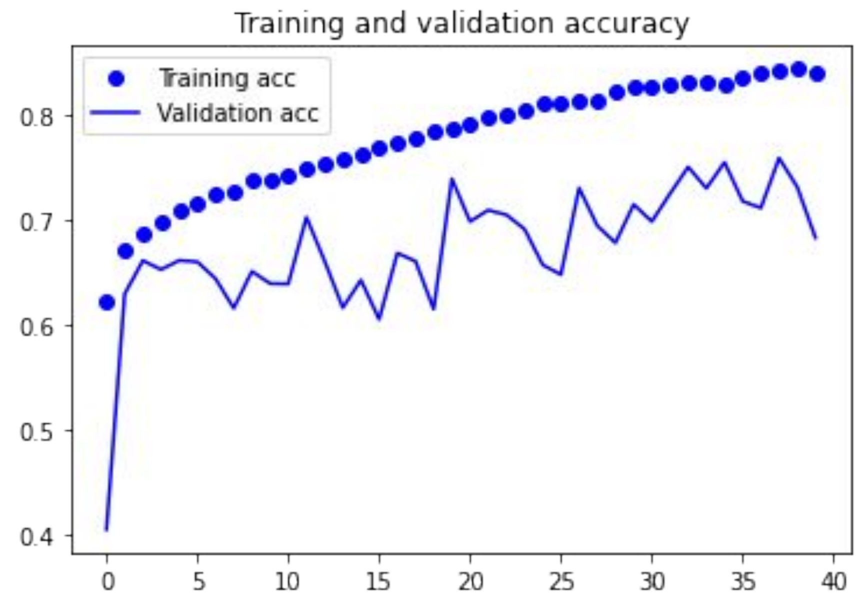


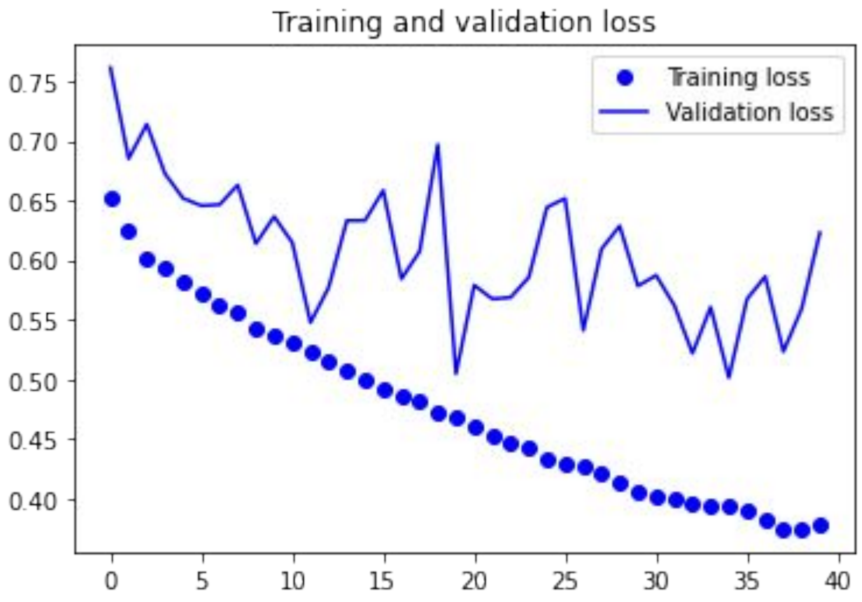




Results: GRU(set B, 48hrs) (2)







Conclusion: GRU

Table 1: The AUROC for Mortality Prediction and Benchmark

	1-day after	In-hospital
Set A, 24 hrs	0.79	0.81 (0.85*)
Set B, 24 hrs	0.80	**0.82** (0.87*)
Set A, 48 hrs	0.77	0.78 (0.86*)
Set B, 48 hrs	0.78	0.79 (0.86*)

Table 2: The AUPRC for Mortality Prediction and Benchmark

	1-day after	In-hospital
Set A, 24 hrs	0.092	**0.343**
Set B, 24 hrs	0.085	**0.342**
Set A, 48 hrs	0.055	0.294
Set B, 48 hrs	0.058	0.301

Conclusion:

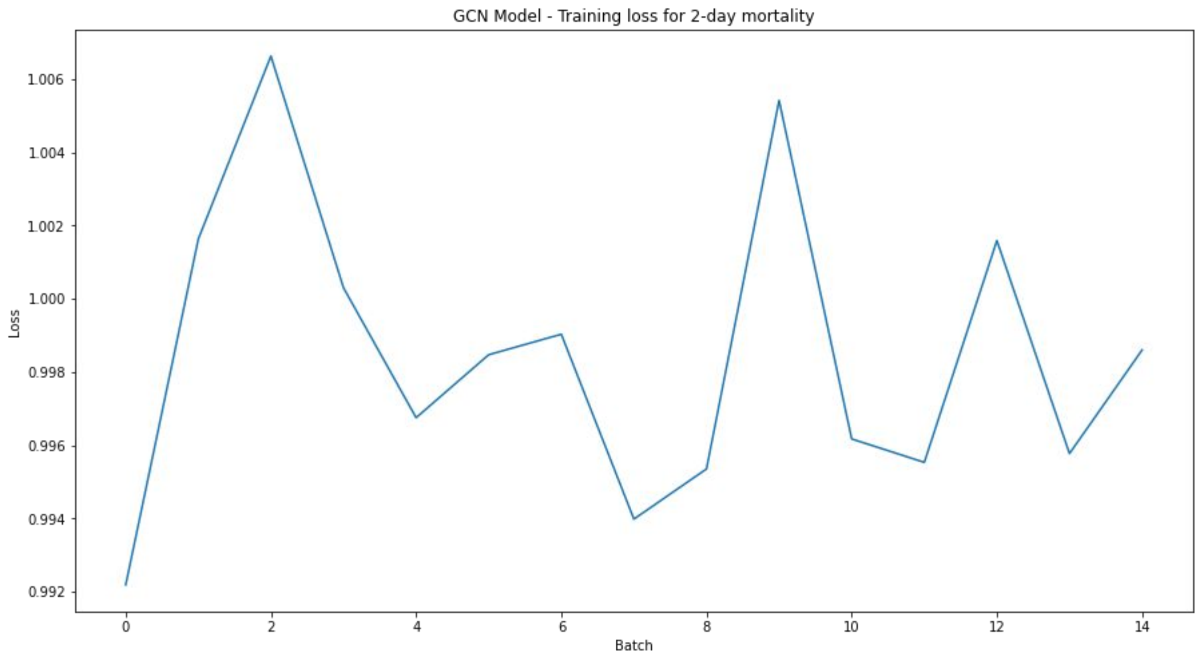
- Overall, GRU gives a better prediction on **in-hospital mortality** (less skewed)
- Based on AUROC, Set B (using raw features) **outperforms** Set A (using processed features) in all cases

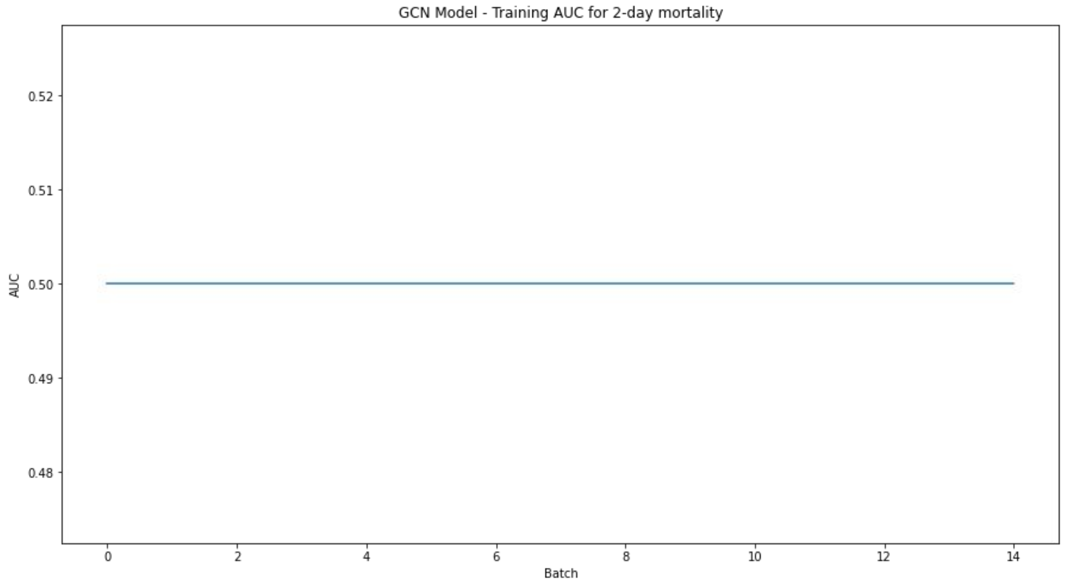
Improvements:

- More complex architecture (ensemble of FFN and GRU) can achieve better results, up to 0.85-0.87*
- Try to add more raw features: 99 else can be processed

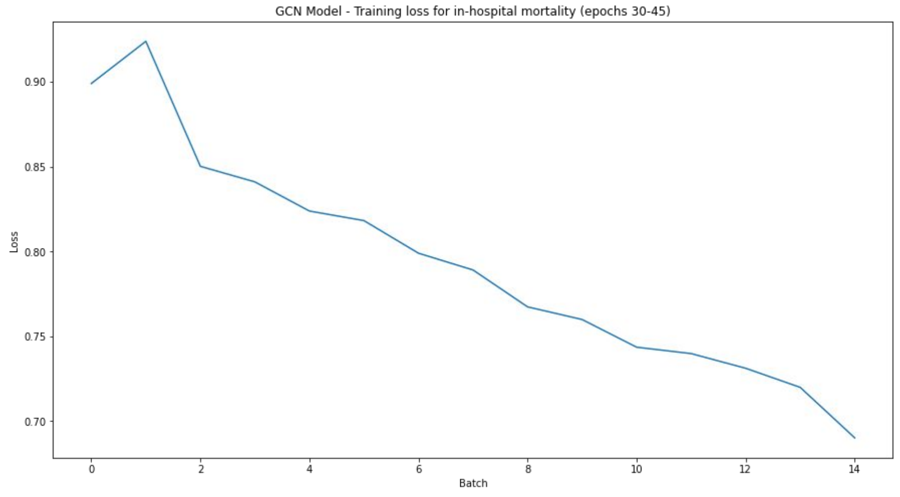
** source: Sanjay Purushotham, Chuizheng Meng, Zhengping Che, and Yan Liu. Benchmark of deep learning models on large healthcare mimic datasets. arXiv preprint arXiv:1710.08531, 2017.*

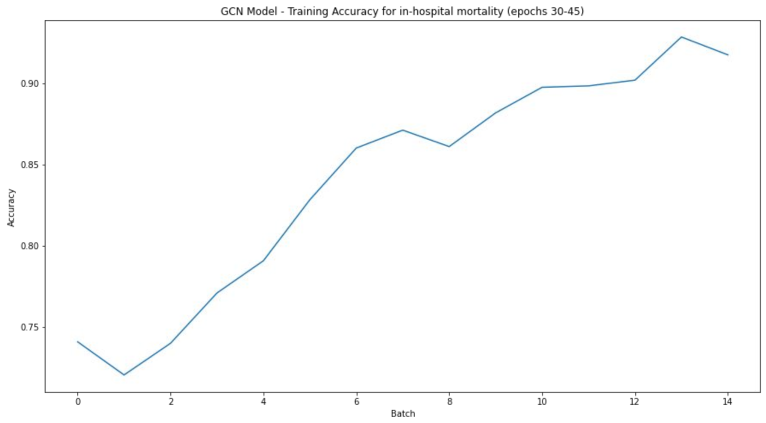
Results: GNN(set A, 24hrs) (1)

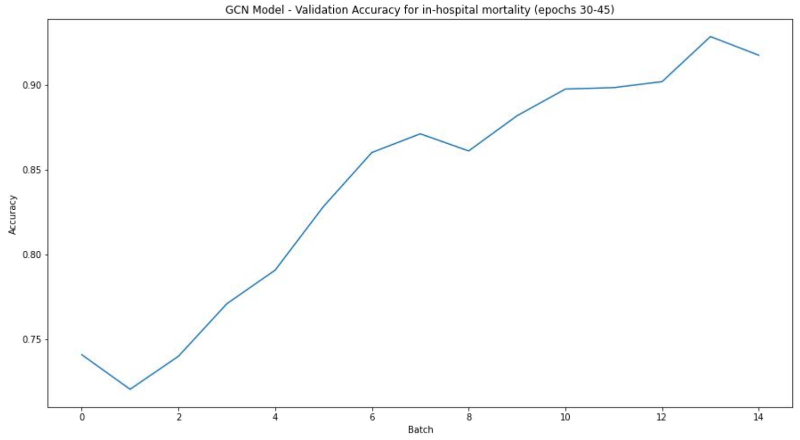


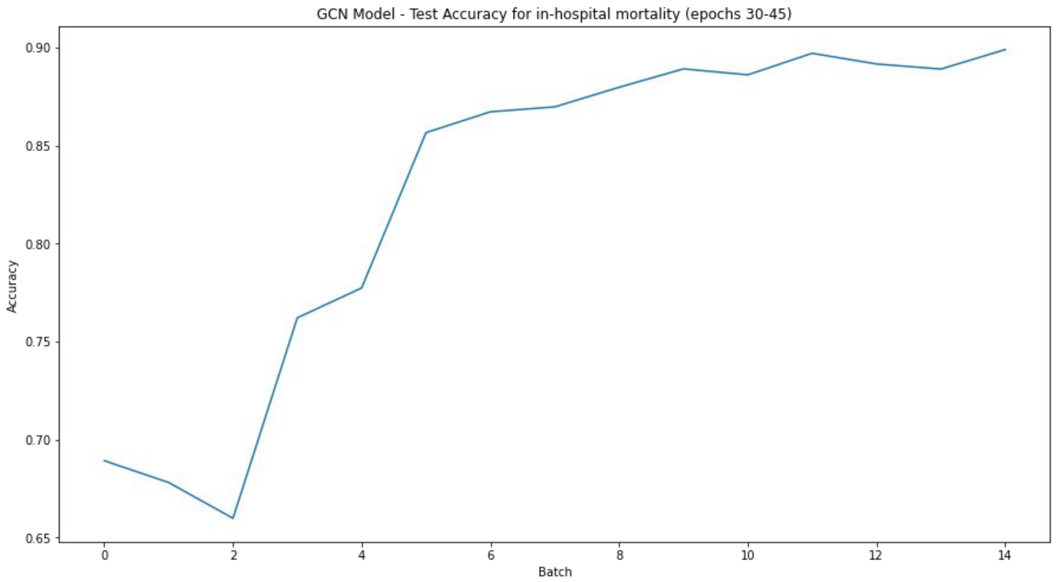


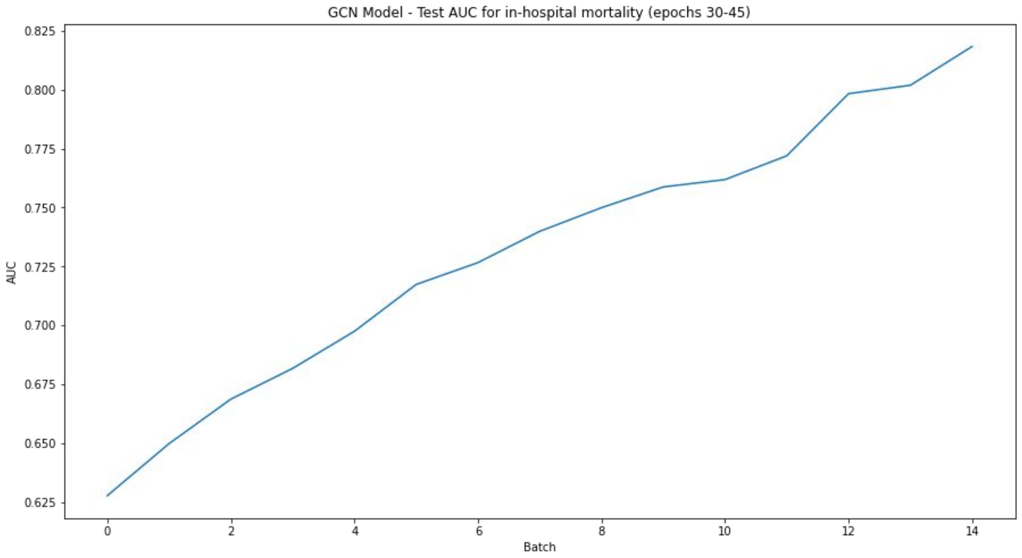
Results: GNN(set A, 24hrs) (2)











Conclusion: GCN

Table 1: The AUROC for Mortality Prediction and Benchmark

	2-day after	1-day after	In-hospital
Set A, 24 hrs	0.50	0.50	~0.82

Conclusion:

- Overall, GCN also works better for **in-hospital mortality** (less skewed) than 1 or 2 day mortality.

Future Experiments:

- Further fine-tuning is possible
- Test for Feature Set B
- Different types of GCN layers can be experimented with (eg. GATs).

Teamwork:

- **Literature Review** – Yuan & Poornima
- **Data Preprocessing / Feature Extraction** - Yuan & Poornima
- **GRU implementation** – Yuan
- **GNN implementation** – Poornima
- **Final presentation / Report** – Yuan & Poornima

References:

1. Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-wei, Mengling Feng, MohammadGhassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. Mimic-iii, afreely accessible critical care database.Scientific data, 3:160035, 2016.
2. Sanjay Purushotham, Chuizheng Meng, Zhengping Che, and Yan Liu. Benchmark of deep learningmodels on large healthcare mimic datasets.arXiv preprint arXiv:1710.08531, 2017.
3. Weicheng Zhu and Narges Razavian. Graph neural network on electronic health records for predictingalzheimer’s disease.arXiv preprint arXiv:1912.03761, 2019.