Neural LoS
A neural way to measure Length of Stay

Priyank Jain, Vibhor Jain

Why Length of Stay (LoS)?

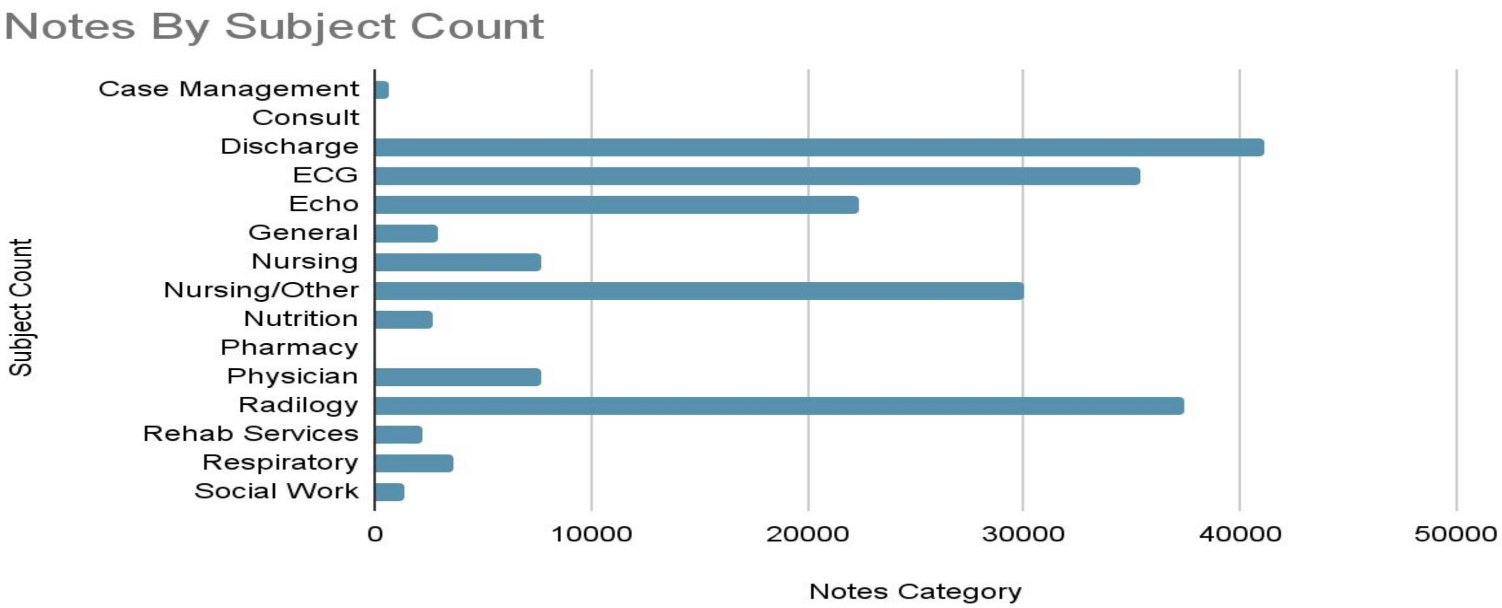
Per the research done by us below are points that paved way to select this topic

- LoS is one of the important areas in medical field, where predicting might help in
  - Planning readmission rates
  - Predicting mortality rates
  - Cost Control
- Clinical Notes plays a very important role in predicting LoS
- Very few papers discussed about merging physiological Data with clinical Notes
- Those few papers which used the idea employed traditional ML like logistic regression

Data Analysis and Preprocessing

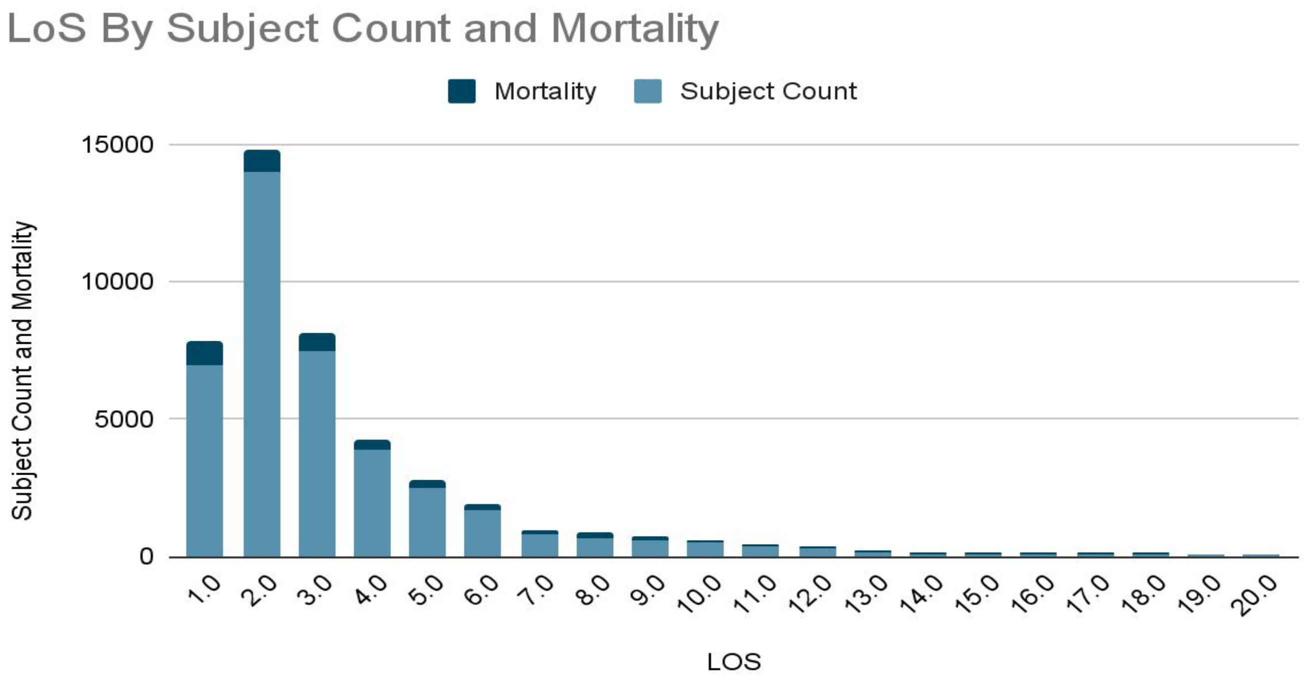
- Preprocessing of physiological data is already available
- Hard to come up with pre processing of notes
  - Multiple ways experimented
- Format and type of each column needs to be determined depending on discretization
- Lot of data analysis is carried out to come up with the chosen approach

Data Analysis and Preprocessing (Continue..)

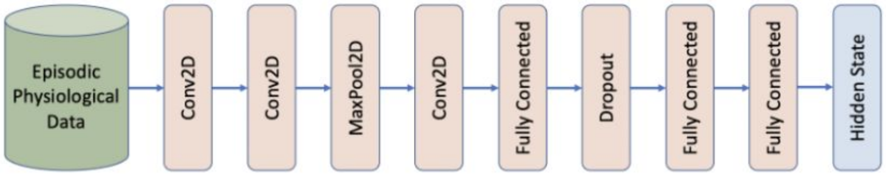


Notes Category vs Subject Count

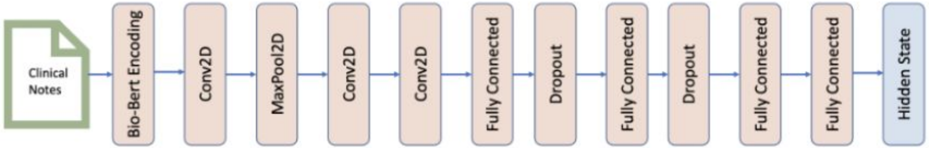
Data Analysis and Preprocessing (Continue..)



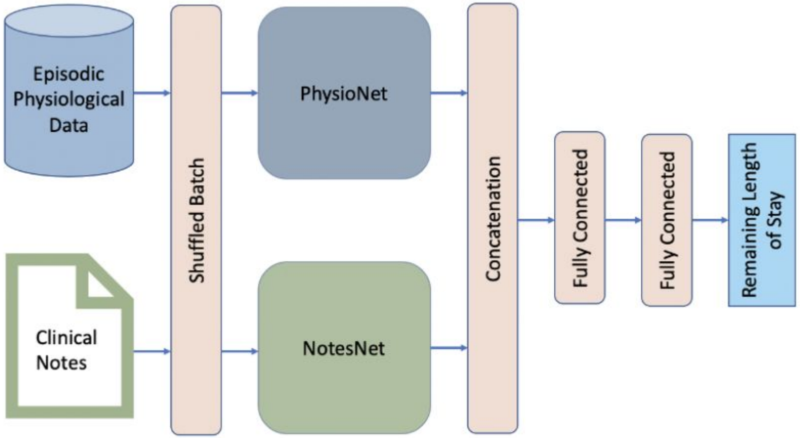
Model Architecture



PhysioNet



NotesNet



EpisodeNet

Methods

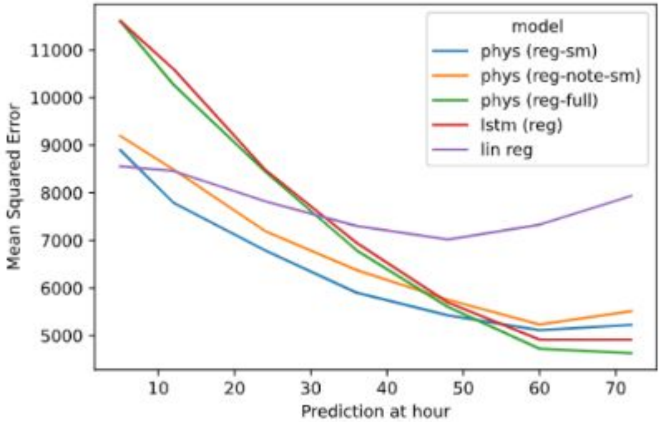
Full details are provided in the paper. Here is the snapshot of few:

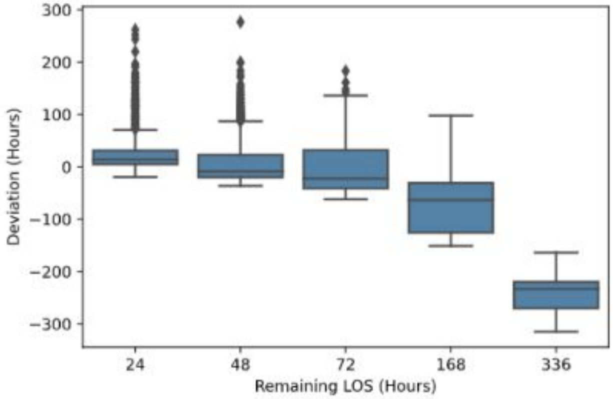
- ❏ Extracting and Cleaning relevant Note Events depending on run capacity in VM
- ❏ Created Note embedding using
  - ❏ BERT Embeddings with self modeling
  - ❏ BioSent2Vec/PubMed using existing model

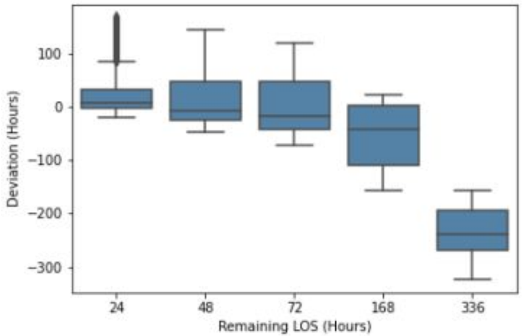
Methods Continued..

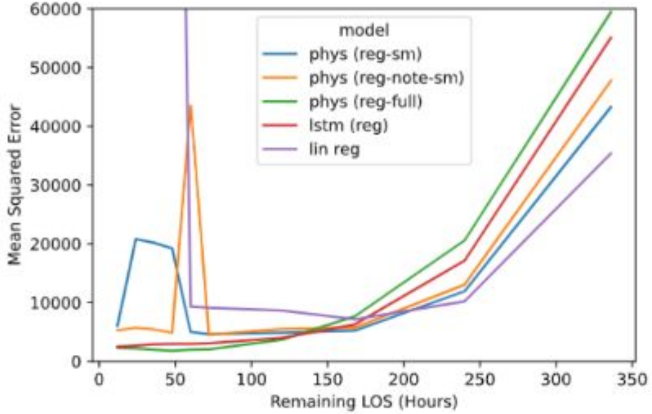
- ❑ Modeling both physiological and Clinical Notes in parallel
- ❑ New approach on Analyzing results
- ❑ LoS Prediction Looking better after combining Clinical notes with Physiological data

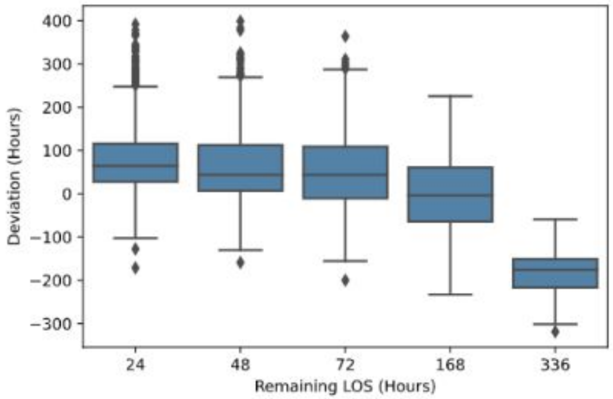
Results


model
phys (reg-sm)
phys (reg-note-sm)
phys (reg-full)
lstm (reg)
lin reg
Mean Squared Error
Prediction at hour


Deviation (Hours)
Remaining LOS (Hours)


Deviation (Hours)
Remaining LOS (Hours)


model
phys (reg-sm)
phys (reg-note-sm)
phys (reg-full)
lstm (reg)
lin reg
Mean Squared Error
Remaining LOS (Hours)


Deviation (Hours)
Remaining LOS (Hours)

Results

Model	Data Types	MAE	MSE	MAPE
Linear regression	Tabular	121.69	12,805,595	3.15
LSTM	Tabular	**79.28**	16,889	81.10
LSTM	Notes	101.44	28,201	**0.72**
PhysioNet (full data)	Tabular	78.55	17,492	1.01
PhysioNet (Part data)	Tabular	80.50	17,450	1.66
PhysioNet + Notes (Part Data)	Tabular	80.49	**16,122**	1.55

Model Results for length of stay prediction

Challenges Faced

We faced several challenges during our project. Some of them are listed below:

- ❑ Choosing the virtual machine with GPU to balance between cost and speed
  - ❑ Finding GPU resources was the challenge in provisioning the instance WITH nvidia Driver AMI GPU driver due to limited capacity of CPU and submitted the request to AWS to increase the capacity from 8 vCPU to 32 vCPU
  - ❑ Eg: BERT took 13 hours to process on V00. BioSent2Vec took 33 min on a single CPU. BERT embeddings take up 54GB of space and BioSent2Vec 40GB

Challenges Faced Continued..

- Identifying feasible ways of pre processing the data
  - Approach and Format of data for pre processing
  - Lack of time to create a model from data (e.g. PubMed/BioSent2Vec)
  - Idea of combining both physiological and text at preprocessing stage and come up with one model was put on hold
  - Issues with notes not available in the initial hours as we can not have default values like physiological data
  - Restricted Notes to 40 sentences per henry and 40 words per sentence to finish training in reasonable time y selecting only Nursing/Other and radiology categories of Text

Future Work

- ❏ Add demographic information as a feature for the subject (Patient)
- ❏ Adding a sequential relationship between episodes and attention
  - ❏ Currently each episode is treated independently
- ❏ Due t computation resources limitations, we trained our EpisodeNet model on a subset of data. We would like to see the odel performance when trained on entire dataset

Resources Used

- ❑ GitHub
- ❑ MIMIC-III Database
- ❑ Reference Papers
- ❑ AWS
- ❑ MIRO
- ❑ Overleaf