TATA CONSULTANCY SERVICES

Deciphering Clinical Narratives – Augmented Intelligence for Decision Making in Health Care Sector

Dr. Lipika Dey
TCS Research, India

Keynote Talk

19th September, 2023

FedCSIS
18th Conference on Computer Science and Intelligence Systems
FedCSIS 2023
Warsaw - Poland
17-20 September

Outline of the talk

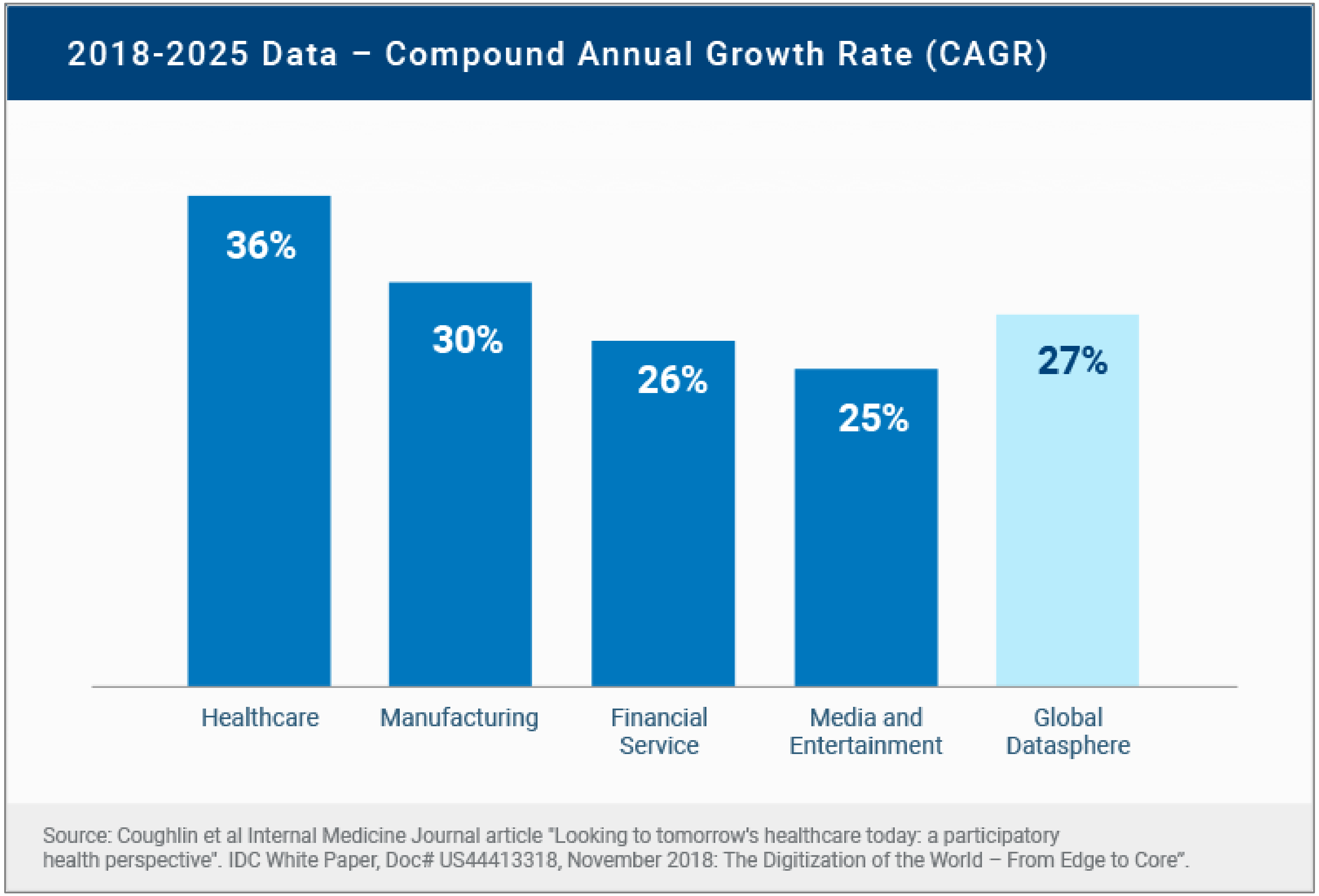
- Clinical texts and Decision-Making problems in Health Care Sector
- Predicting Length of stay in ICU using first day's nursing notes
  - Results and Learnings
- Forming Patient Cohorts
  - Deeper dive into risk factors for each group
- Way forward – some plans

Clinical narratives - Main form of communication within health care

- **Clinical Data**
  - Electronic Medical Records (EMR)/Electronic Health Records (EHR)
  - Physician and Care-giver notes - patient history, assessments and treatments
  - Clinical trials management – trial description, monitoring trial progress
- **Social Media (tweets, Facebook comments, message boards, etc.)**
  - personal accounts of patients – signals for mental health – adverse effects of drugs
  - Health care system feedback
- **Medical Literature**
  - News feeds, Medical journals
- **Insurance Providers (claims from private and government payers)**
  - Underwriter notes

Was estimated to be 25,000 petabytes by the end of 2020 – COVID 19 enhanced it by many orders

Most rapidly rising data repository



https://www.rbccm.com/en/gib/healthcare/episode/the_healthcare_data_explosion

Challenges of working with Clinical Data

Chances of Privacy violation

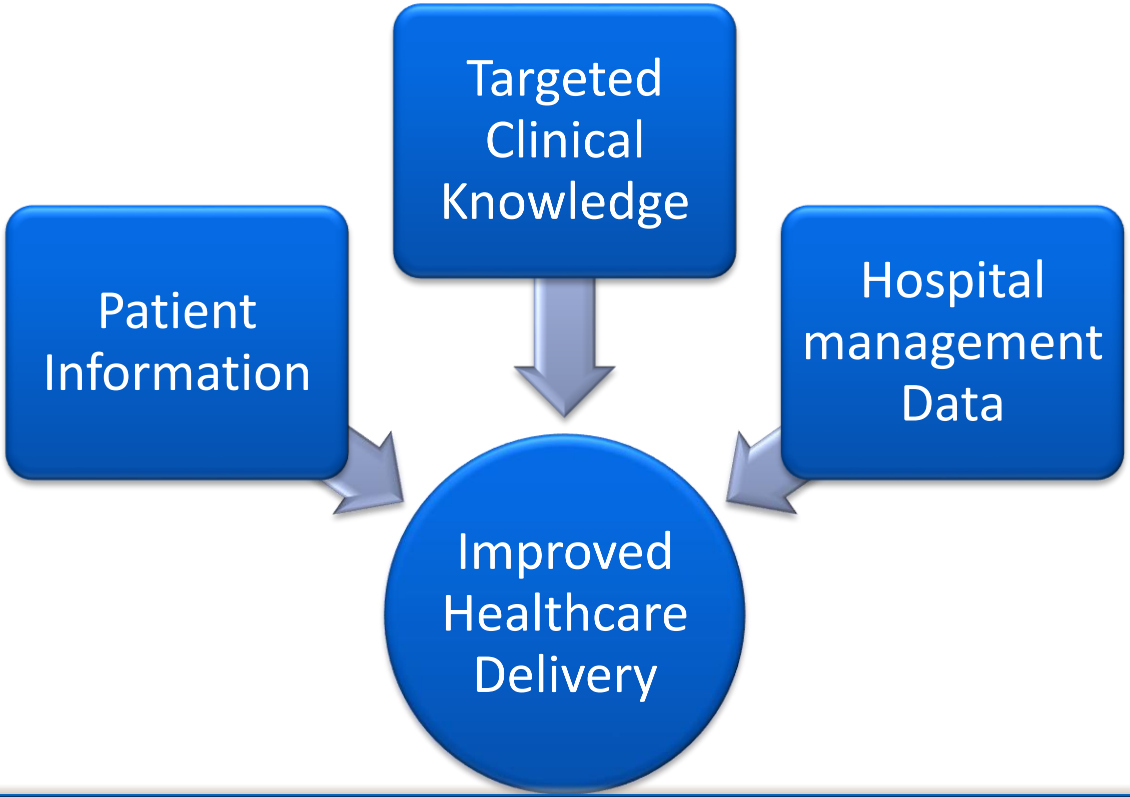
Not available in large quantities for research

Makes training and evaluation of Machine Learning techniques difficult

Non-standard terminologies

Textual narratives are more so

Decision Support Systems in Health Care Sector



- Hospital Logistics Management
- Expectation management for patients and their family
- Personalized Patient Care

Using Clinical Notes

ICU LENGTH OF STAY PREDICTION

About MIMIC-III v1.4 Dataset

- MIMIC-III v1.4 Database contains details of 58976 admission records of 46,520 patients who stayed in critical care units of the Beth Israel Deaconess Medical Center (BIDMC) between 2001 and 2012 - developed by the Laboratory for Computational Physiology, MIT.
- Has pre-existing ***Institutional Review Board (IRB) approval***
- Adheres to ***stringent anonymization protocols*** - meticulously safeguards patient privacy
- Ensures heightened privacy protection by ***obfuscating precise dates and times of events***
- Researchers can access the data ***after successfully completing the training course*** "Data or Specimens Only Research" provided by the Collaborative Institutional Training Initiative (CITI)

ICU Length of Stay (LOS) Prediction – ***why is this important?***

- According to the World Health Organization (WHO) patient Length of Stay (LOS) in hospitals is an important performance measurement and monitoring indicator
- Intensive Care Units (ICU) are premium facilities in a hospital – comprising hospital resources, manpower, and equipment
  - Accurate prediction of patient LOS aids healthcare specialists to take medical decisions and allocate medical team and resources appropriately
- Better logistics planning – ensures ***better resource usage*** for critically ill patients
- Patient and insurance companies may use this prediction to manage their budget

The Earlier the better

Earlier Work – using Patient Health Parameters

Comparing with SOTA	Dataset	Feature used	Method	Best Result
Alghatani et al., 2021	44,000 ICU stays from MIMIC	patient's vital signs like, heart rate, BP, temp., resp. etc	Random Forest	65% accuracy
Su et al., 2021	2224 Sepsis patients PICMISD	Age, P(v-a)CO_2 /C(a-v)O, SO, wbc etc.	XG-Boost model	F1: 0.69, AUC-ROC:0.76
Rocheteau, Liò, et al., 2020	eICU critical care dataset	medical features, Gender, Age, Ethnicity, etc.	Temporal convolution	Kappa score = 0.58
Harutyunyan et al., 2019	42276 ICU stays of 33798 unique patients from mimic database	17 clinical variables like, Capillary refill rate, Diastolic blood pressure etc. from first 24 hours of admission.	LSTM	AUC-ROC : 0.84
van Aken et al., 2021	38013 admission notes from MIMIC III	Created admission notes from ***discharge summaries***	Pretrained CORe + BioBERT	AUC-ROC : 0.72%

Nursing Notes - a complete clinical narrative of hospital admission

The pt is a complicated 62yo man who was transferred from ajh last evening with mrsa bacteremia and pnx. Pt arrived via EMS,intubated, sedated on Propofol 15mcg/kg-min, on Dopa gtt at 5mcg/kg-min. Transferred to Big Boy bed with 6-person assist, MICU-A monitor, and MICU-A IV pumps.**Dopamine** titrated up to max of 6.5mcg/kg-min with SBP 90's-80's, Propfol weaned from 15mcg/kg-min -> 12. Levo-phed started at 0.05mcg/kg-min and dopa weaned to 4.0mcg at time of shift report. Propofol d/c'd, and fentanyl and midazolam started at 25mcg/kg-min and 0.5mg/hr, respectively......Pt turned upon admission; SBP by a-line dropped to 70's with + wave-form, and SpO2 dropped to 80's**.....Skin breakdown noted** over back of neck, sloughed skin with serosanguinous drng OTA on arrival......NaHCO3 3 amps given after 2nd ABG when acidosis was worsening with repsiratory intervention....Daughter in to see pt. Gravity of pt status discussed with daughter.....Patient remains on mechanical ventilation;switched to PCV due to high Paw.PIP improved as well as Pa02,but patient still has significant metabolic acidosis.**Renal insufficiency,patient may need to be dialysised.BS diminished,suctioned for small amount of clear thick tenacious type of secretion. the micu team is concerned that the pt may have mrsa endocarditis and the plan is to obtain a tte later today.** he continued to have a difficult noc with persistent fevers, hypoxia, acidosis, copd/emphysema, asbesteosis, endometriosis with hematuria as well as requiring an increase in pressor support.....

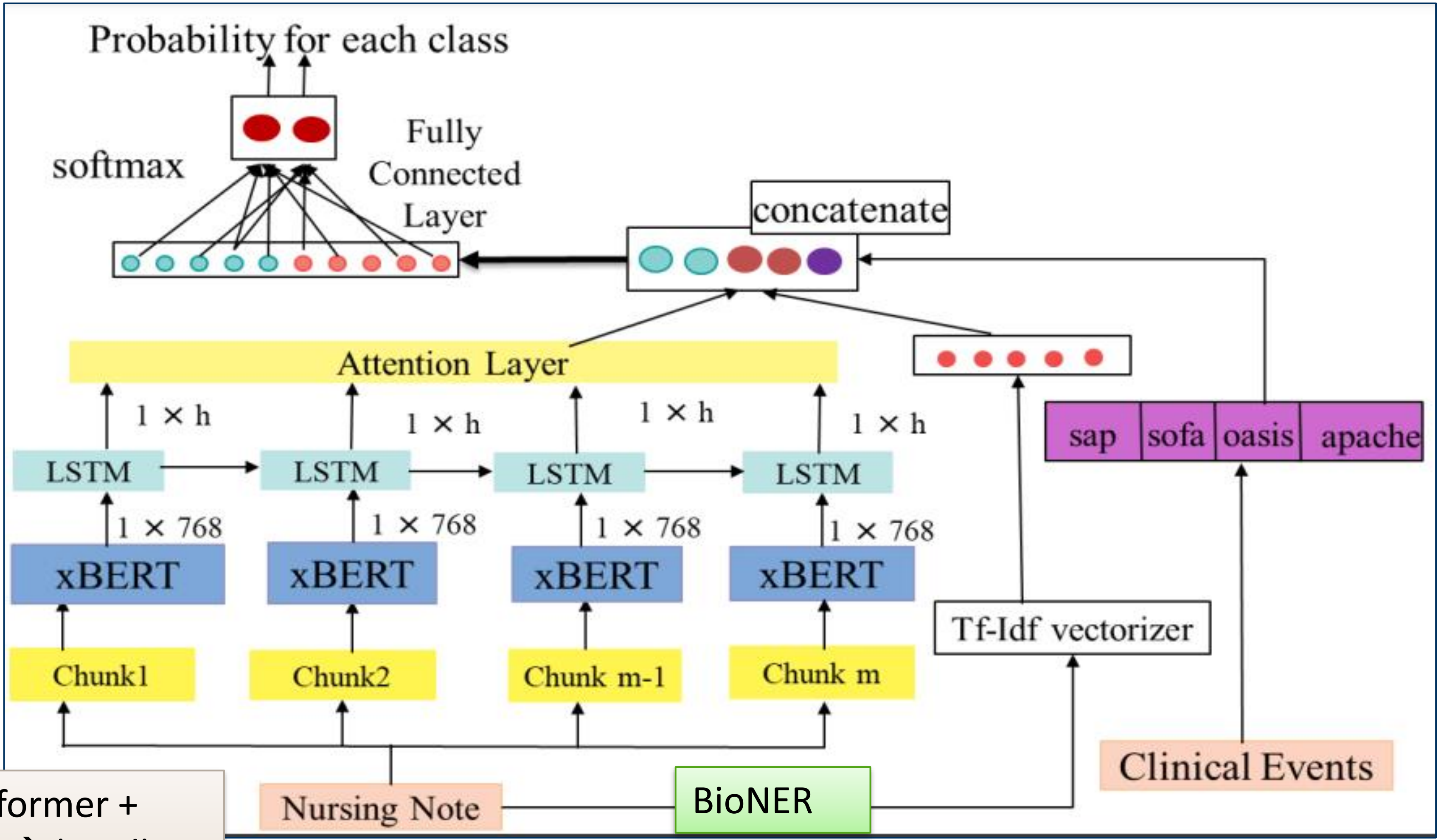
Nursing notes are very detailed, time-stamped account of a patient's stay in a hospital

- Observations made by Doctors about symptoms and diseases
- Prior medical history including drug allergies etc.
- Critical examinations suggested and / or reports
- Treatment plans
    - Medicines prescribed
    - Procedures suggested
- General conditions – appetite, mobility, pain

Predicting ICU stay on the day of admission to ICU

- Most of the earlier work used vital parameters to predict LOS
- Van Aken et al., 2021 used Discharge Summaries - contains entire history of patient during admission
- **Our work**
  - ***Use only the first day's Nursing Note prepared after patient’s admission to ICU to predict the LOS***
- Though called LOS –
  - usually prediction is about LONG / SHORT stay – where SHORT <= MEDIAN, LONG > MEDIAN
  - Exact stay is decided based on multiple extraneous factors like age, availability of facility, patient willingness etc.

Multimodal DNN for Predicting ICU LOS



Transformer + LSTM → handles long notes

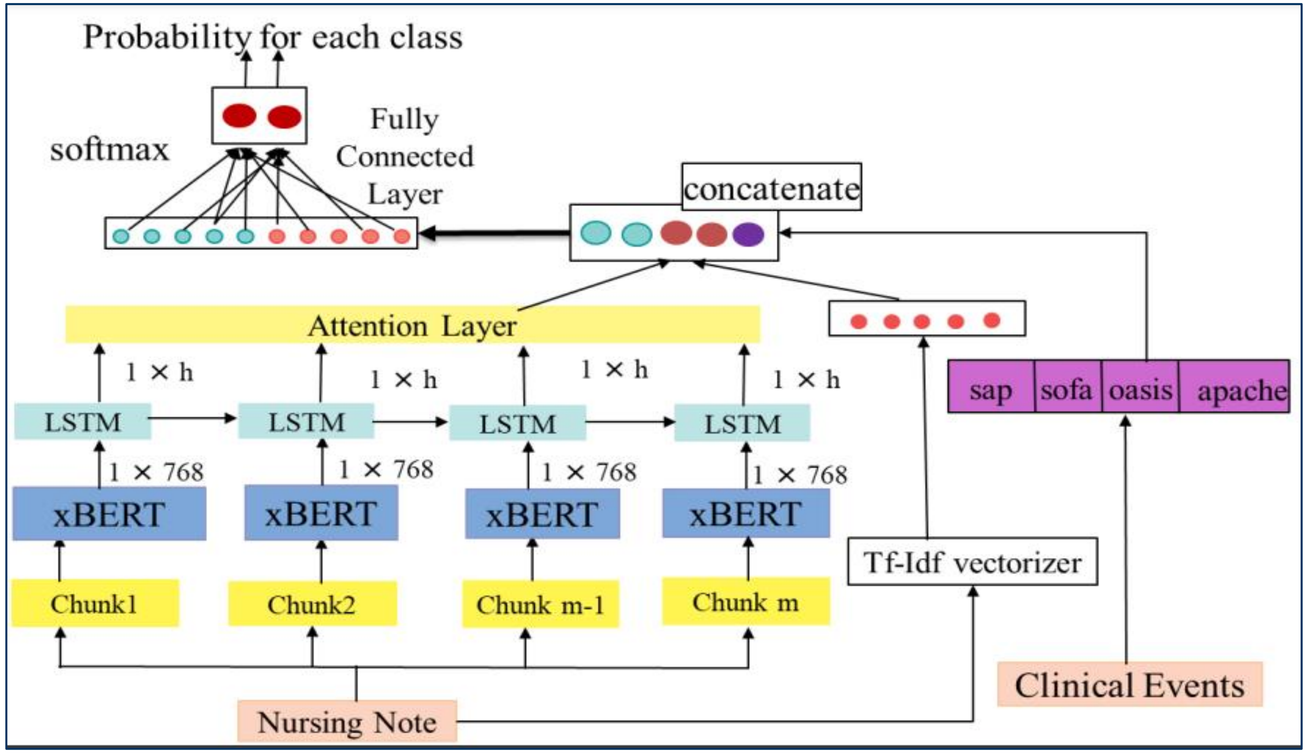
Tf-IDF of Linguistic features extracted using BioNER (CLINER, SciSpacy) - Clinical Named Entities like drugs, diseases, treatments

Clinical Events

Used *libraries* to obtain Severity of Illness Scores – computed from physiological parameters

- Acute Physiology and Chronic Health Evaluation (APACHE-II) score
- Simplified Acute Physiology Score (SAPS-II)
- Sepsis- related Organ Failure Assessment (SOFA) score
- Oxford Acute Severity of Illness Score (OASIS)

LIME Framework to add explainability



xBERT = BlueBERT and Clinical BioBERT

Local Interpretable Model-agnostic Explanations

Attempts to understand the model by perturbing the input of data samples and observe how the predictions change

- *Weighted average from multi-head attention models*

Output

Long ICU Stays - "HR dropping", "requiring mask ventilation for resp. failure", "couldn't breathe"

Short ICU Stays - "good effect from Ativan", "comfortable breathing", "hemodynamically stable"

Results obtained from 22789 admissions which had Nursing Notes for first day

Model	Accuracy	F1 score Class 'Short'	F1 score Class 'Long'	AUC-ROC
BlueBERT+LSTM+Attn+TF-IDF+SOI	**0.797**	**0.810**	**0.790**	0.872
BlueBERT+LSTM+Attn+TF-IDF	0.792	0.800	**0.790**	**0.873**
BlueBERT+CNN+TF-IDF	0.789	0.800	0.780	0.872
BlueBERT	0.776	0.790	0.760	0.833
Clinical BioBERT+LSTM+TF-IDF+SOI	0.780	0.790	0.770	0.871
Clinical BioBERT+LSTM+TF-IDF	0.778	0.780	0.770	0.857
Clinical BioBERT+CNN+TF-IDF	0.776	0.780	0.770	0.864
Clinical BioBERT	0.741	0.750	0.730	0.818

Studying the Discharge Notes to gain insights about wrong predictions

- Patients developed complications mid-way

- Complications could be about other organs and not related to the disease on admission

- Initial Nursing Notes not always indicative of type of procedures patients underwent later
  - ***Admitted for Pneumonia – underwent Bypass Surgeries***
  - ***Delay in scheduling unplanned procedure***

- Four most common procedures
  - Bypass Surgery
  - Stenting
  - Tracheotomy
  - Cholecystectomy

Reformulated Problem – Predict LOS and Major Procedure requirements

- ***Helps in better planning of resources***
- ***Better expectation management for patient’s family***

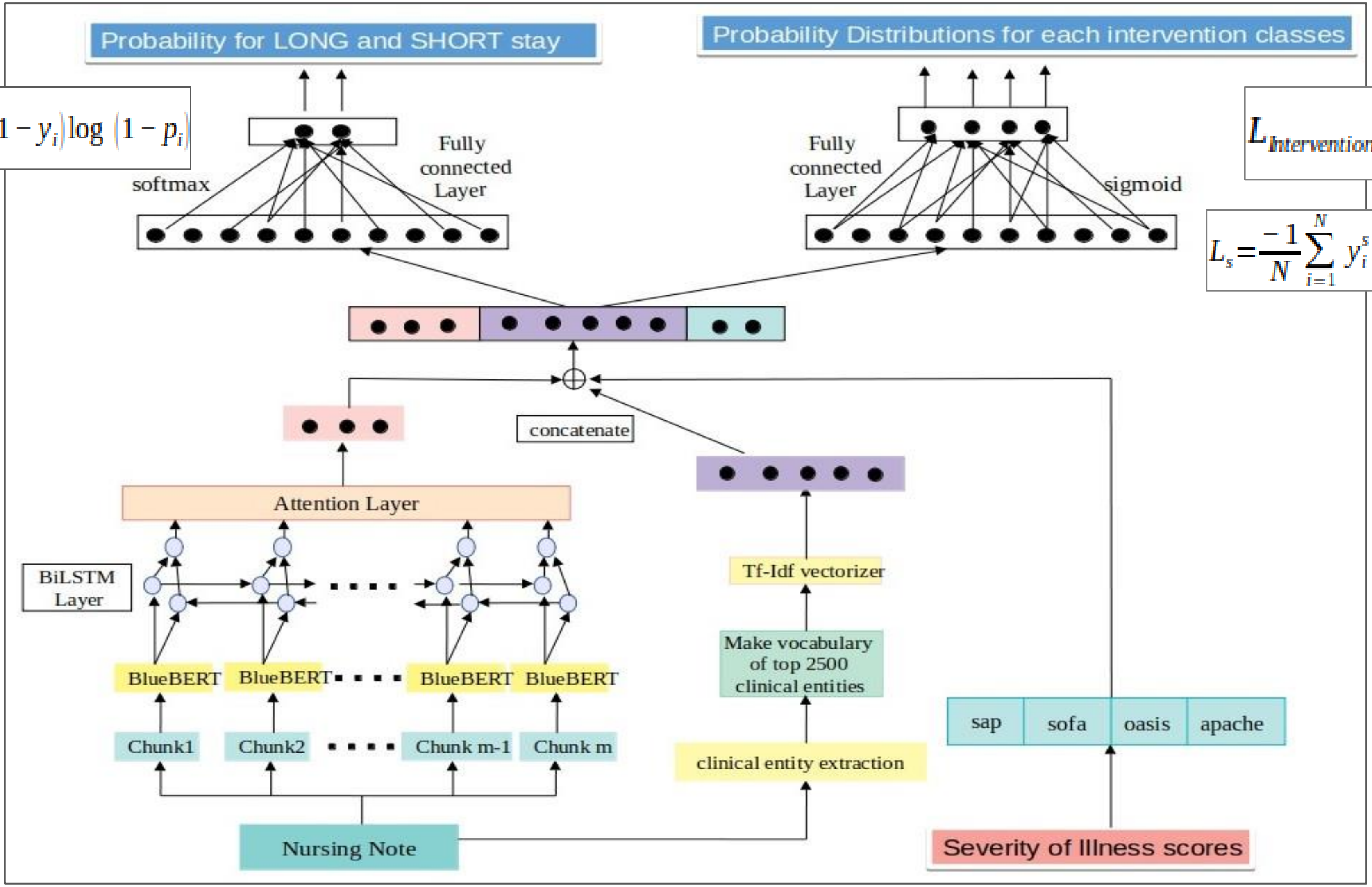
Multitask, Multimodal Architecture for predicting LOS and Major Procedures

$$L_{joint} = \lambda * L_{LOS} + (1 - \lambda) * L_{Intervention}$$

$$L_{LOS} = \frac{-1}{N} \sum_{i=1}^{N} y_i \log p_i + (1 - y_i) \log (1 - p_i)$$

$$L_{Intervention} = \sum_{s \in interventions} L_s$$

$$L_s = \frac{-1}{N} \sum_{i=1}^{N} y_i^s \log p_i^s + (1 - y_i^s) \log (1 - p_i^s)$$



TATA CONSULTANCY SERVICES

Results

ICU LOS Prediction

Model	Accuracy	F1 score 'Short'	F1 score 'Long'	AUC
Multitask BiLSTM-blueBERT with tf_idf and SOI	**0.84**	**0.86**	**0.82**	**0.89**
Multitask BiLSTM-blueBERT with tf_idf	0.83	0.84	0.80	0.87
Multitask BiLSTM-blueBERT	0.79	0.79	0.78	0.85
BiLSTM-blueBERT with tf_idf and SOI	0.79	0.81	0.79	0.86
BiLSTM-blueBERT with tf_idf	0.79	0.80	0.78	0.84
BiLSTM-blueBERT	0.77	0.77	0.76	0.83

Procedure Prediction

Model	Accuracy	F1 score "Bypass"	F1 score "Stent"	F1 score "Tracheotomy"	F1 score "Cholecystectomy"	AUC
Multitask BiLSTM-blueBERT with tf_idf and SOI	**0.82**	**0.89**	**0.83**	**0.55**	**0.54**	**0.86**
Multitask BiLSTM-blueBERT with tf_idf	0.81	0.86	0.81	0.53	0.51	0.85
Multitask BiLSTM-blueBERT	0.80	0.85	0.78	0.51	0.48	0.83
BiLSTM-blueBERT with tf_idf and SOI	0.80	0.85	0.79	0.49	0.49	0.83
BiLSTM-blueBERT with tf_idf	0.78	0.81	0.78	0.48	0.46	0.83
BiLSTM-blueBERT	0.77	0.81	0.79	0.47	0.46	0.81

AUC: area under the receiver operating characteristic; SOI: severity of illness.
We report accuracies, AUC scores of the model, and F1 scores of all four classes. Bold values indicate the best performance of our experiments.

Utility of Prediction Model

Procedure Name	Recall	Precision	Gain = (TP-Mentioned_in_firstnote)/total_number)%
Bypass Surgery	0.98	0.82	**47.97**
Stenting	0.75	0.92	**31.92**
Tracheotomy	0.71	0.44	**68.15**
Cholecystectomy	0.41	0.78	**30.93**

Model was able to predict many procedures on first day itself – ***doctors did it much later***

Identifying cohorts within Patient Data

OBJECTIVE:

BETTER RISK ASSESSMENT FOR EACH COHORT

BETTER PREDICTION OF RESOURCES

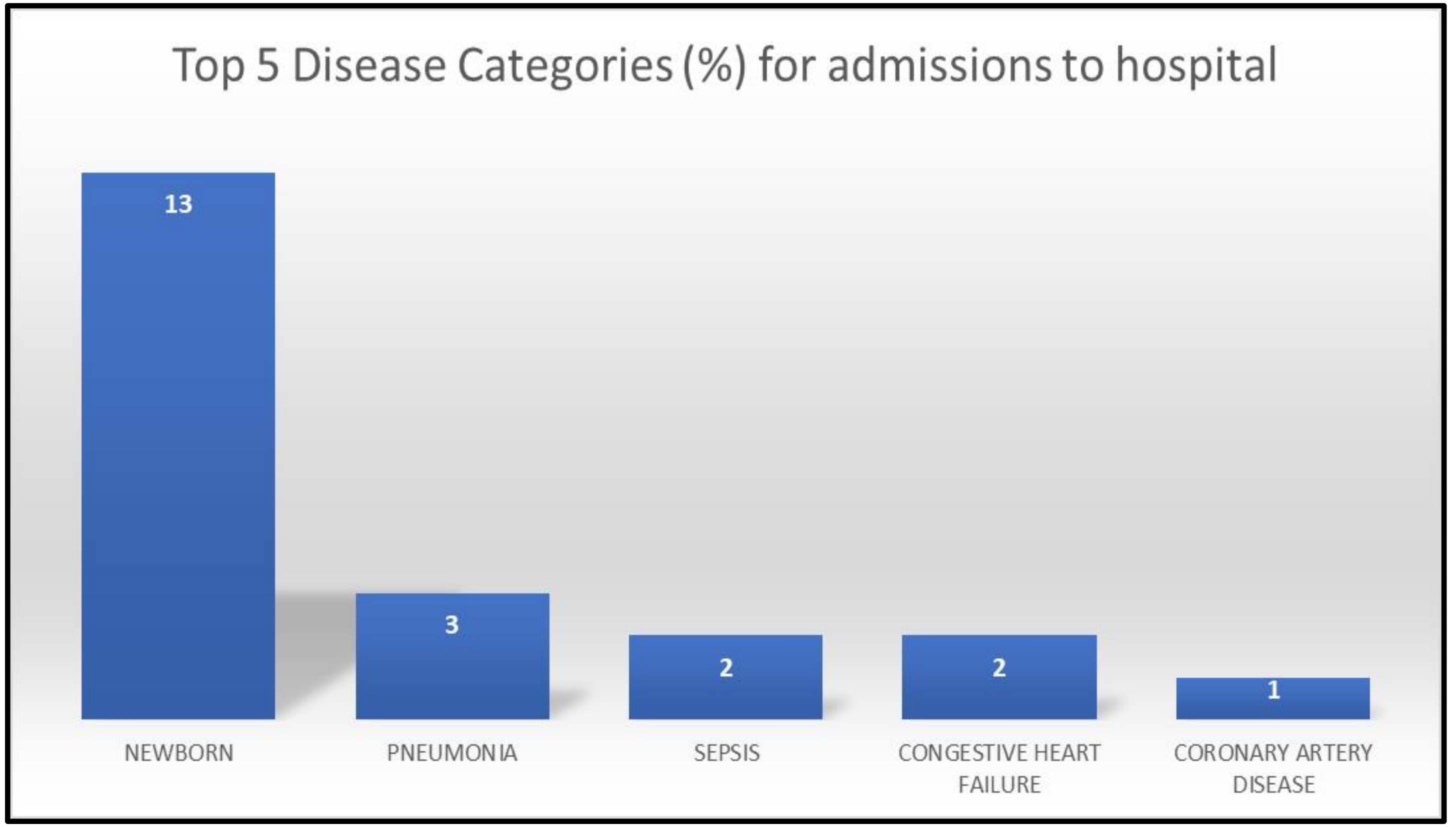
Cohort Discovery for analyzing patient data

Clustering Nursing Notes – observations and learnings

- Clustering Nursing Notes directly did not yield good results
  - ***Use of Bio-medical Resources can help***

- Working with entire dataset for cohort detection did not yield meaningful insights
  - ***Cohort identification is usually done on subsets of patients admitted for a particular disease***

- Nursing Notes of NewBorn is noisy – contains information about mother and neo-natal



Chose Pneumonia as a subset for experimentation

Processing Biomedical Entities using Metathesaurus

Using SciSpacy and Metamap

Scispacy detects Disease names Metamap detects

- Disease or Syndrome
- Acquired Abnormality
- Congenital Abnormality
- Mental or Behavioral Dysfunction
- Injury or Poisoning
- Mental process
- Anatomical Abnormality
- Sign or Symptom

Detect Negation and its argument using Negex

Example
patient denied chest pain → neg (chest pain)

Standardization using Unified Medical Language System (UMLS) - Entity Resolution

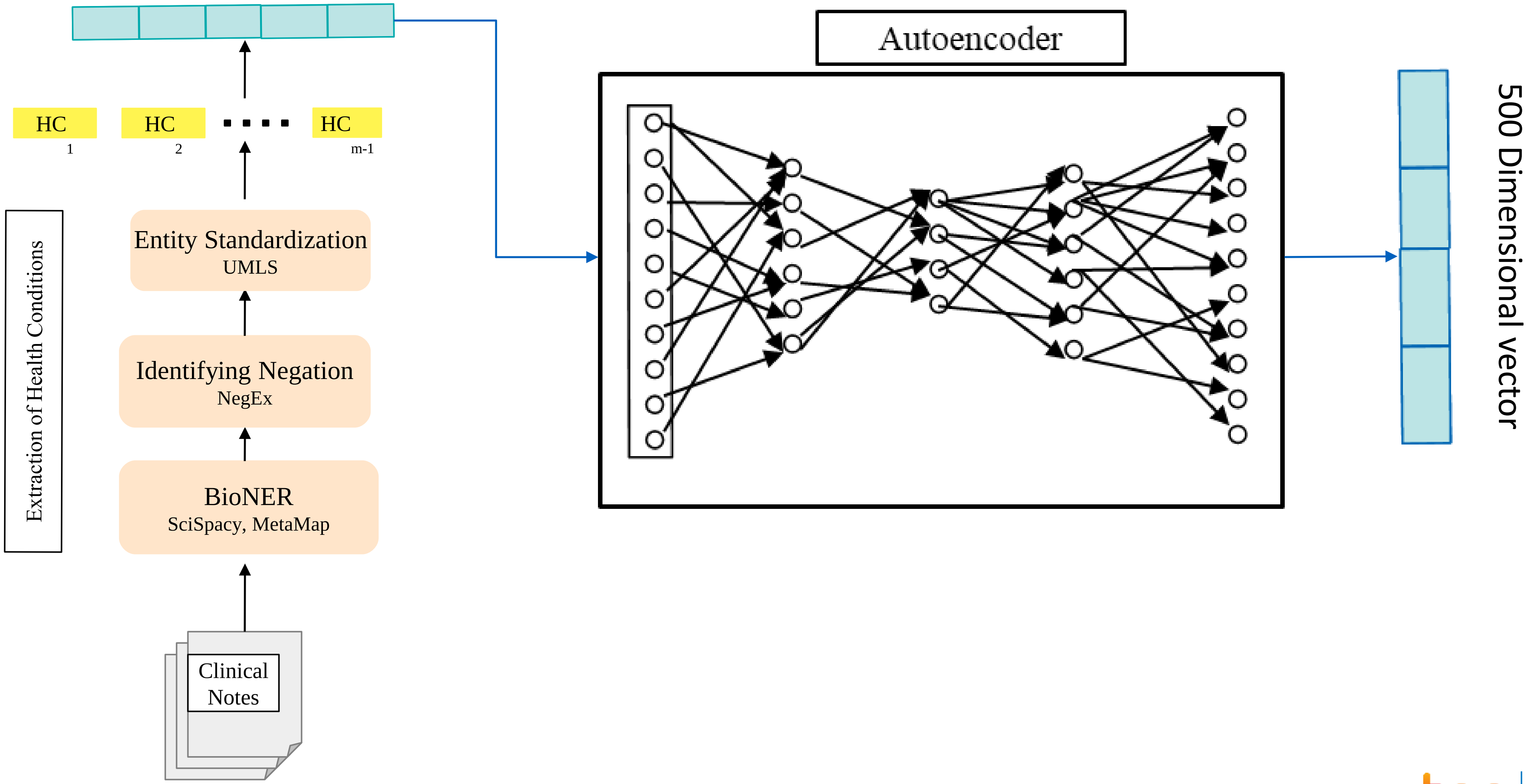
Example - Hypertension, High Blood pressure, Hypertensive Disorder, Arterial Hypertension

Union of all entities used as Basis for representation

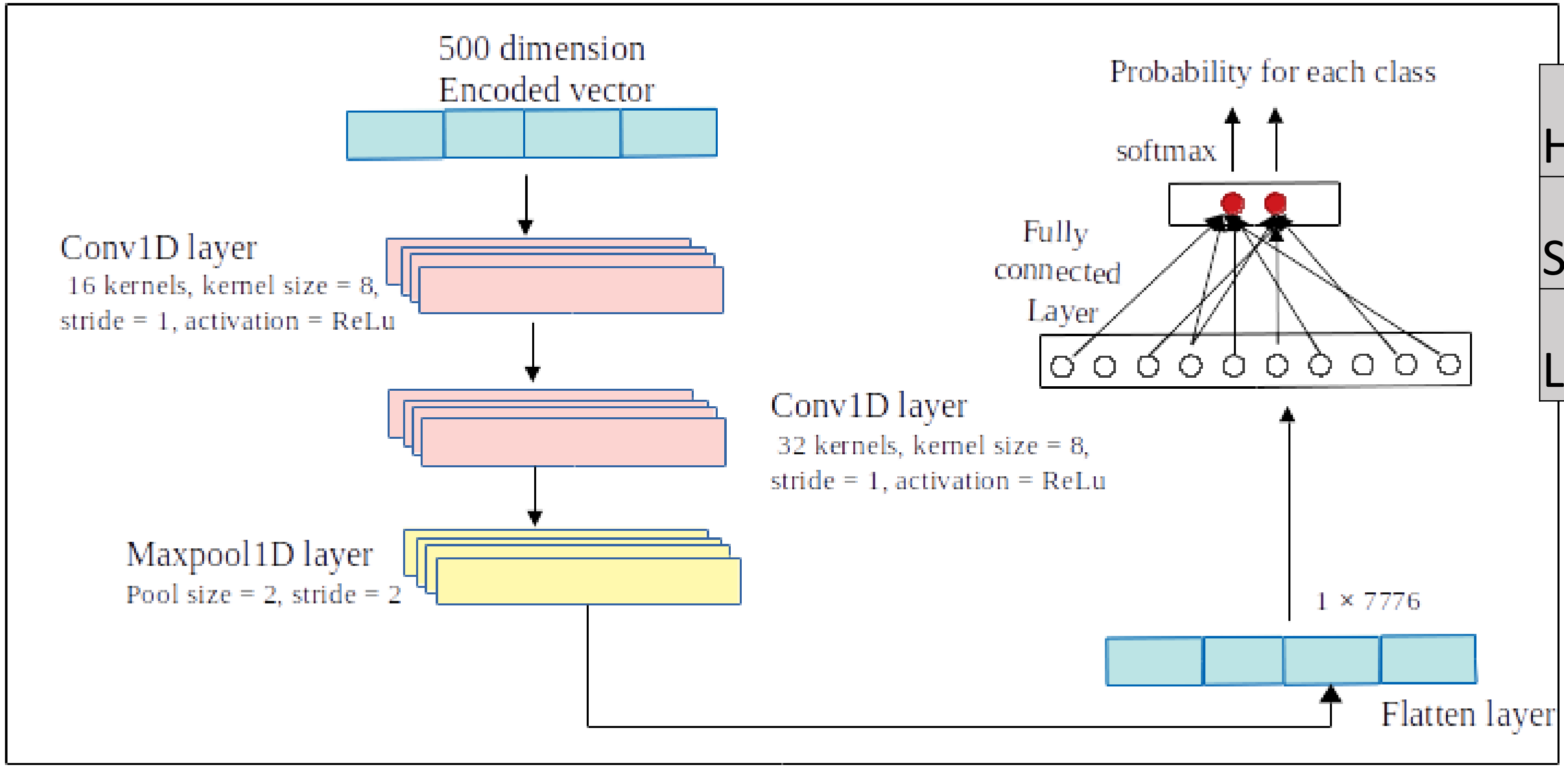
4000+ unique entities

Autoencoders for representation of Nursing Notes

Binary vector of 4541 health conditions – Large and Sparse



LOS Prediction for Pneumonia Patients using Autoencoders



Hospital Stay	Recall	Precision	F1-score
Short	0.64	0.94	**0.76**
Long	0.97	0.80	**0.88**

Accuracy Score : 0.83

K-means Clustering of Auto encoded Nursing Notes

- Distance Metric - Euclidean distance
- Set ***k = 2***
- Iterate to find right value of ***k*** by optimizing the Silhouette Score
  - Measure Silhouette score of each point ***i*** belonging to C_I as

$$s'(i) = \frac{b'(i) - a'(i)}{\max\{a'(i), b'(i)\}}$$

  - where $a'(i) = d(i, \mu_{C_I})$ $\quad b'(i) = \min_{C_J \neq C_I} d(i, \mu_{C_J})$
  - Compute Silhouette Coefficient as

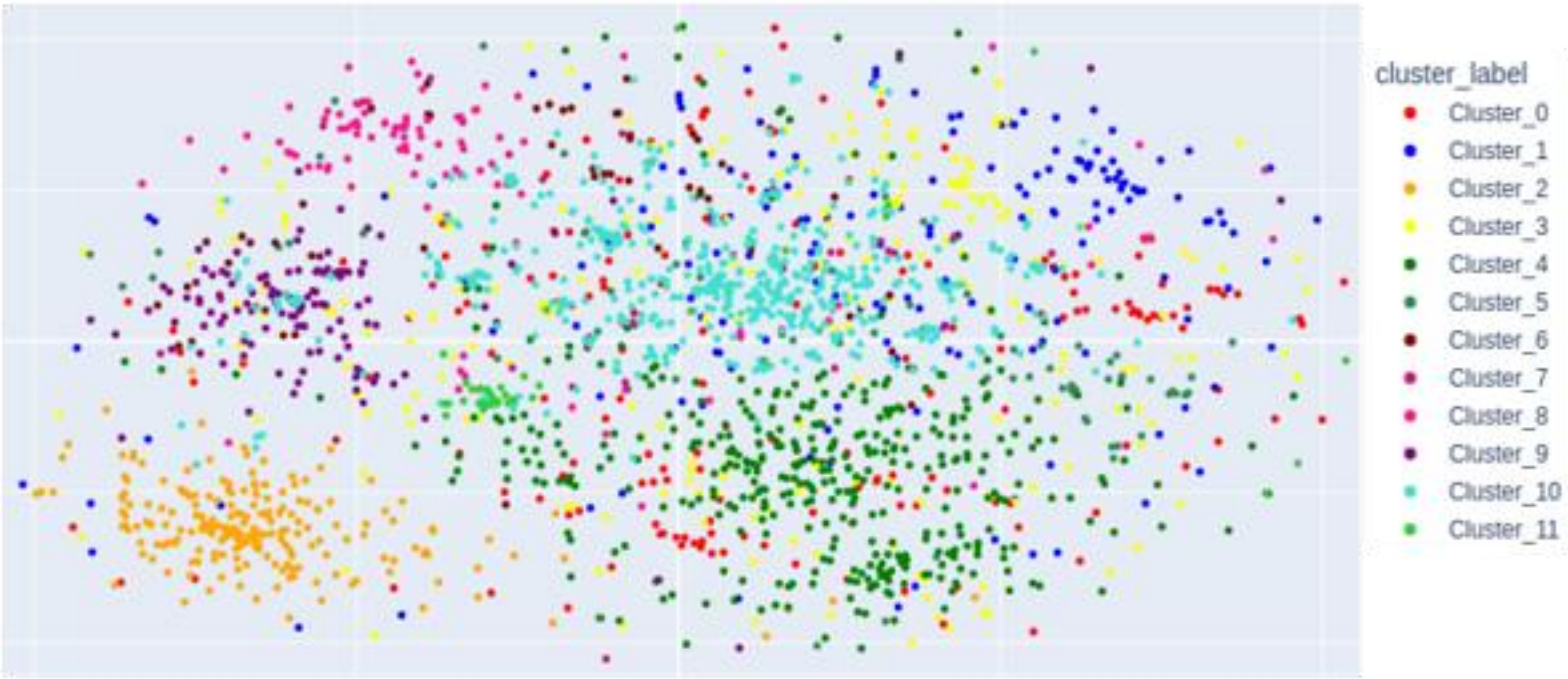
$$SC' = \max_k \frac{1}{N} \sum_i s'(i)$$

- Find ***k which minimizes*** Silhouette Coefficient

Silhouette coefficient of each point measures its relative distance from its own cluster center and other centers - ***cohesiveness versus distinctiveness***

Silhouette coefficient measures the maximum value of the mean of all scores

Clustering 2106 Pneumonia Patients – 2D visualization using tSNE



How to interpret the clusters?

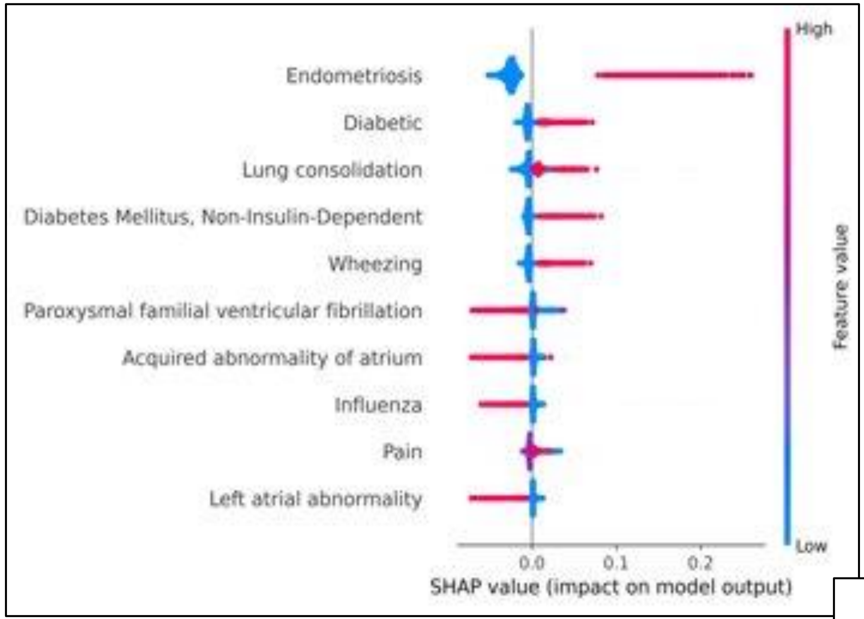
SHAP Values for interpreting clusters

- SHAP values are based on game theory
  - Determines the importance of each feature towards label assignment
- Features with positive SHAP values positively impact the prediction, while those with negative values have a negative impact
- Magnitude measures the strength of impact

- Using the Cluster IDs as label
- The original 4541 entity vector was used as input to a Random Forest Classifier
- Models were fed to SHAP tree explainer
- Obtained the SHAP values for each feature towards determination of the labels

SHAP Explainability for a few clusters

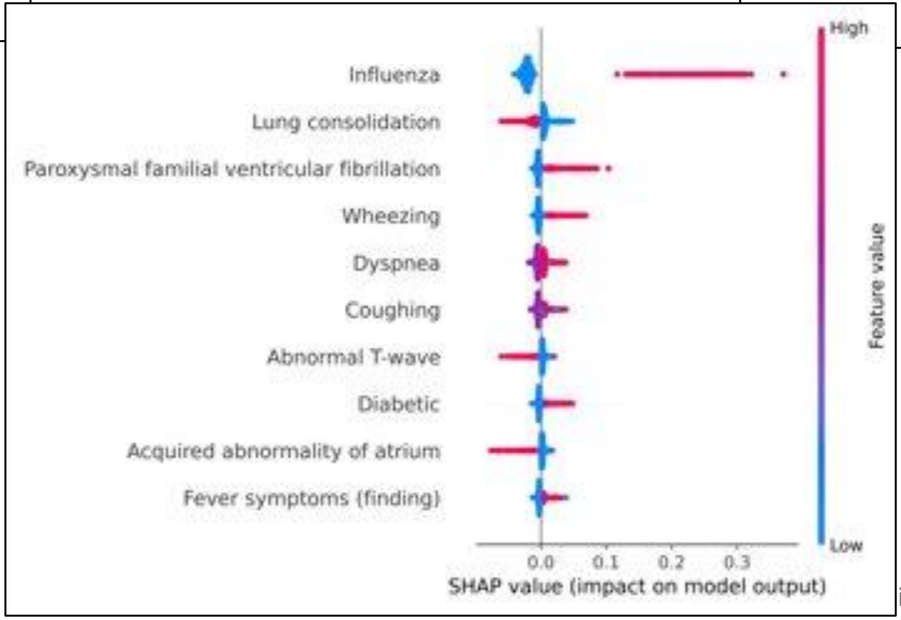
Cluster 2 – Acquired Abnormality of Atrium, Left Atrial Abnormality

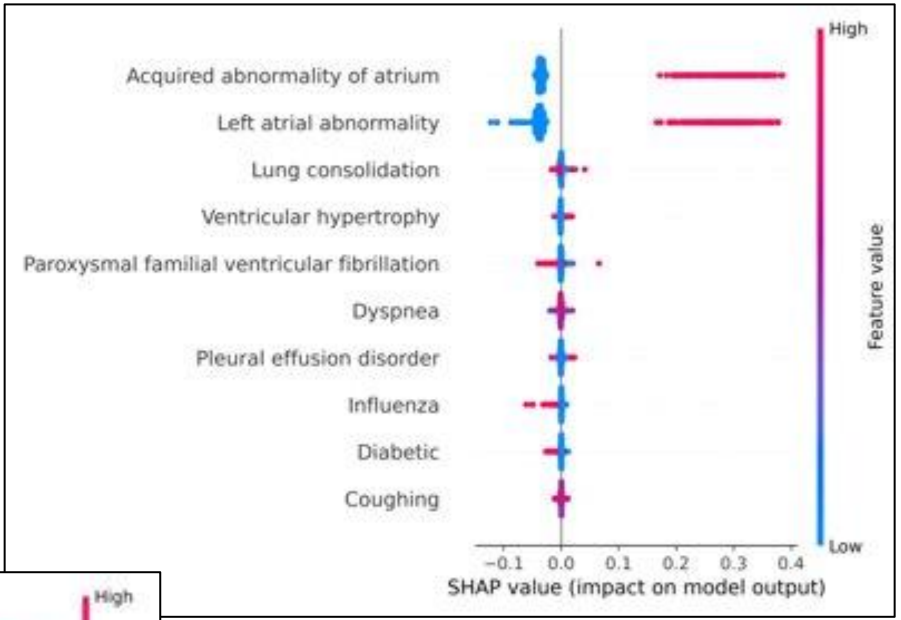


Cluster 0 – Diabetic Patients with Endometriosis

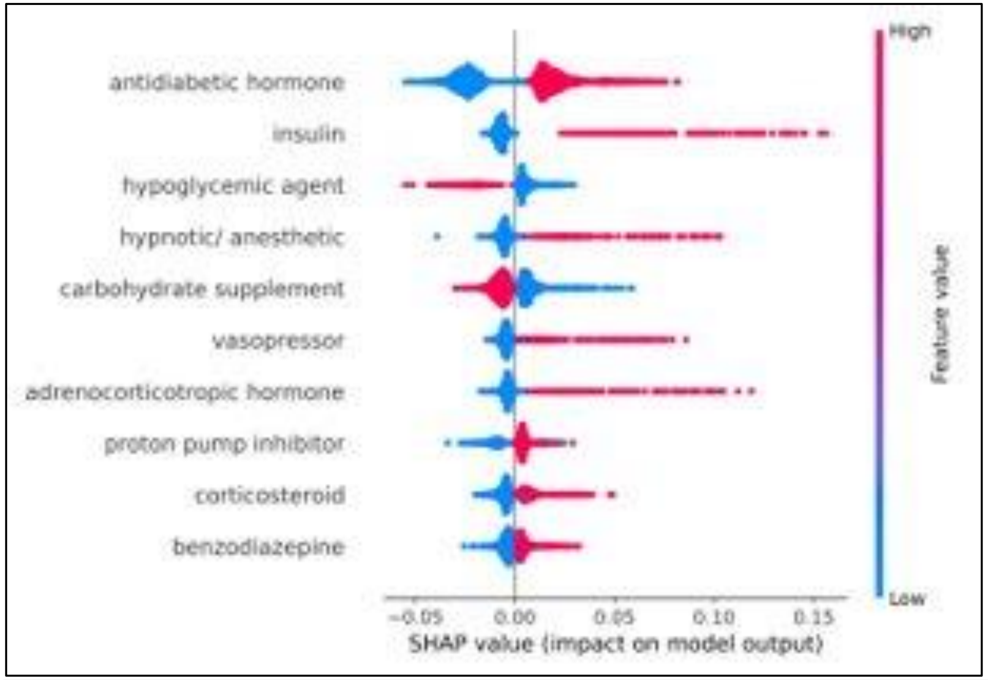
Each cluster could be described uniquely using top three disease / symptom as attributes (present / absent)

Cluster 1 – Influenza, Lung Consolidation absent



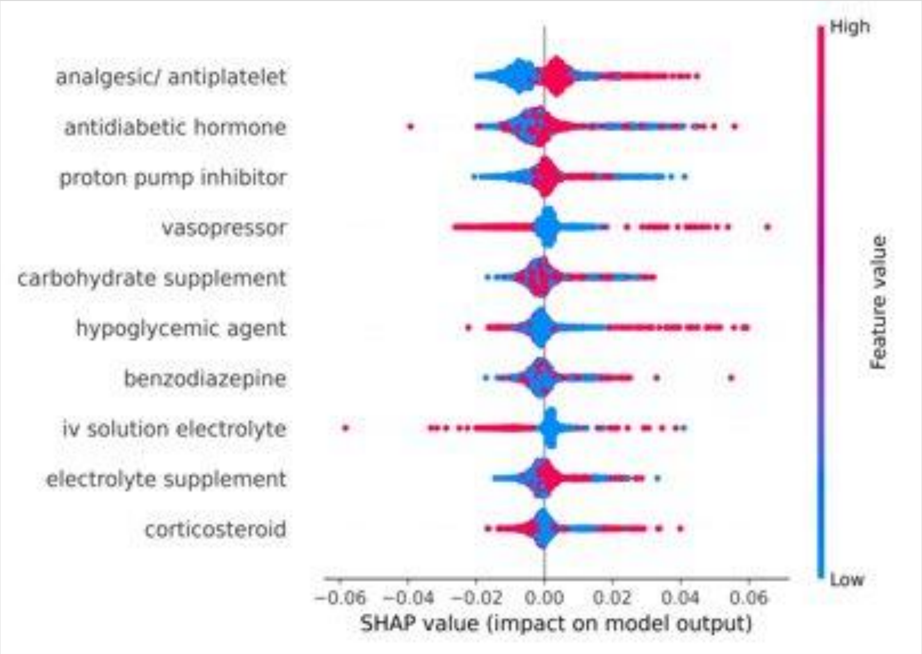


SHAP values for Drug Associations with each cluster



Cluster 0 – Anti diabetic hormone, Insulin, vasopressor

Drug – Cluster Associations were not significantly unique



Cluster 2 – Analgesic, proton pump inhibitor

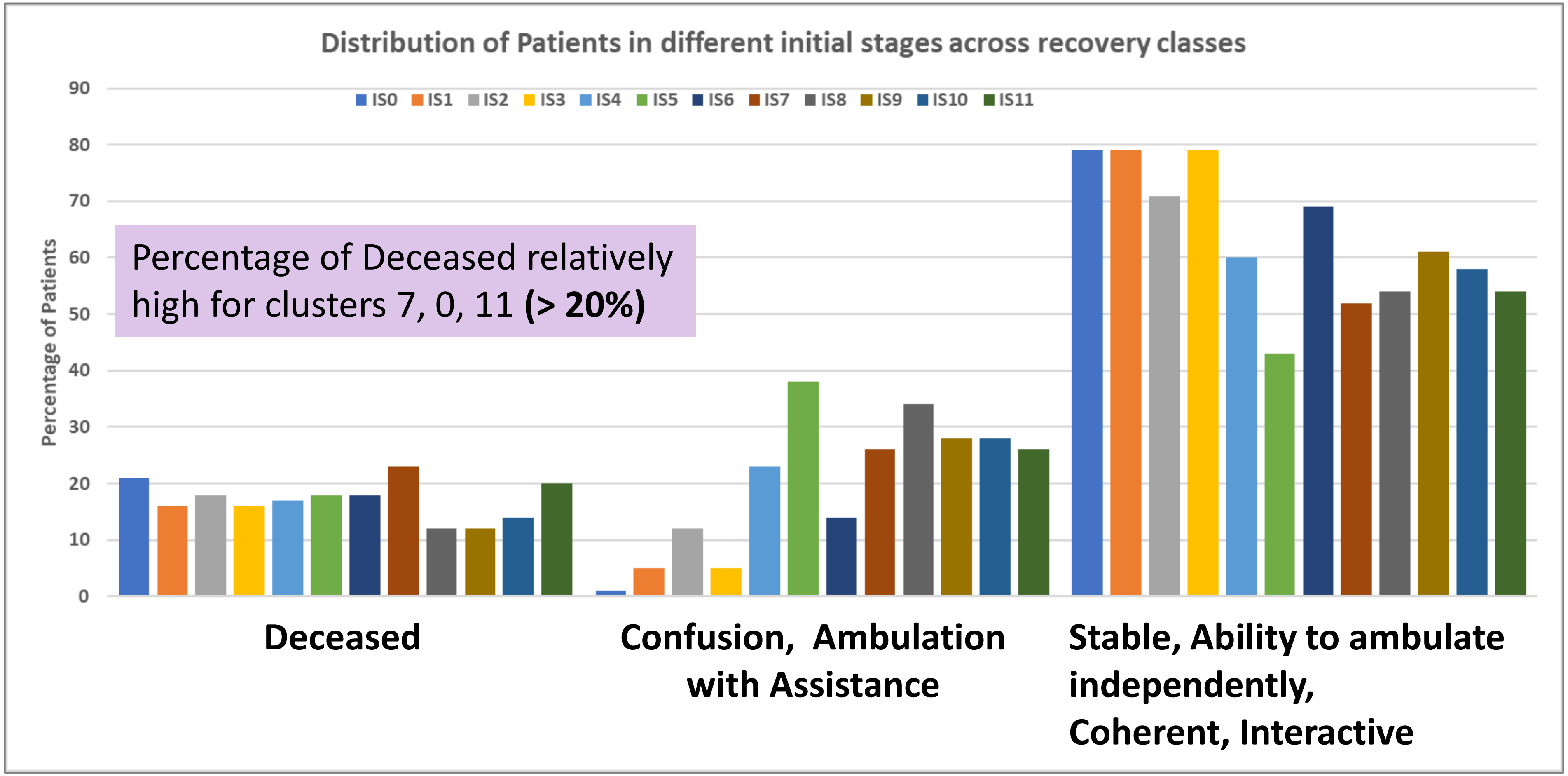
Causality Analysis between Drugs and LOS was further investigated using DoWhy

Disease and Symptoms of each patient were given as attributes – LOS was target variable – drugs as interventions

Example Observation - For patients suffering from ***arterial premature complexes in cluster 6*** - drugs effective in reducing stay were *Corticosteroid* (-3.17), *Anticonvulsant/ neuropathic pain agent* (-3.02), *Opioid analgesic* (-2.4), *Vasopressor* (-1.02)

But significant causal inferences could not be drawn for all clusters - aligned with earlier observation

Cluster-wise Recovery Status (identified from Discharge Notes)



High Risk Pneumonia Patients
Cluster 7 – *Edema, Hypotensive, Left axis deviation;*
Cluster 0 – *Endometriosis and Diabetes;*
Cluster 11 – *Ventricular hypertrophy*

Risk calculation – Probability of Discharge State given Initial Symptoms

P(Discharge State/Symptom) - computed cluster-wise and over whole dataset

End-state is mostly dependent on Comorbidities rather than majority symptoms – rare events and combinations are more informative

- Sample Findings
  - Chronic multifocal osteomyelitis → Deceased (0.85) (*cooccurs mostly with Endometriosis*)
  - Cardiac Arrest → Deceased (0.6)
  - Portal Vein Thrombosis → Deceased (0.6)
  - Renal Osteodystrophy → Deceased (0.6)
  - Diverticulosis of sigmoid colon→ Confusion, Ambulation with Assistance (0.7)
  - ***Most of the above symptoms are distributed across clusters***

To Conclude

- Predictive power of Nursing notes are relatively less explored
  - Can improve prediction of Hospital stay and procedure requirements effectively
  - Accuracy of LOS prediction improves when patients of a single disease are considered
  - Explainable mechanisms can provide insights about predictions
- Better insights are possible to obtain from Patient Cohorts

    - Combination of auto-encoders and SHAP explainability offers rich insights about Risks associated

<u>Work in Progress</u>

- Clustering Mid-stage Symptoms
  - Persistence of symptoms, New symptoms and their associations with recovery stages
- Reconstruction of Recovery Pathways with probabilities
  - Explainable Risk assessment framework
- Multi-modal data from MIMIC Database

Thank You