BERKELEY LAB
Bringing Science Solutions to the World

U.S. DEPARTMENT OF ENERGY

Predicting ICU Readmission Using Context-Enriched Deep Learning

Rafael Zamora-Resendiz

Million Veterans Project Group - Computational Research Division

----

Project Team: Dr. Silvia Crivelli, LBNL

Dr. Xinlian Liu, Hood College and LBNL

Xiange Wang, LBNL

September 26, 2019 -- CompBioMed: Machine Learning, Big Data & AI

Presentation Overview

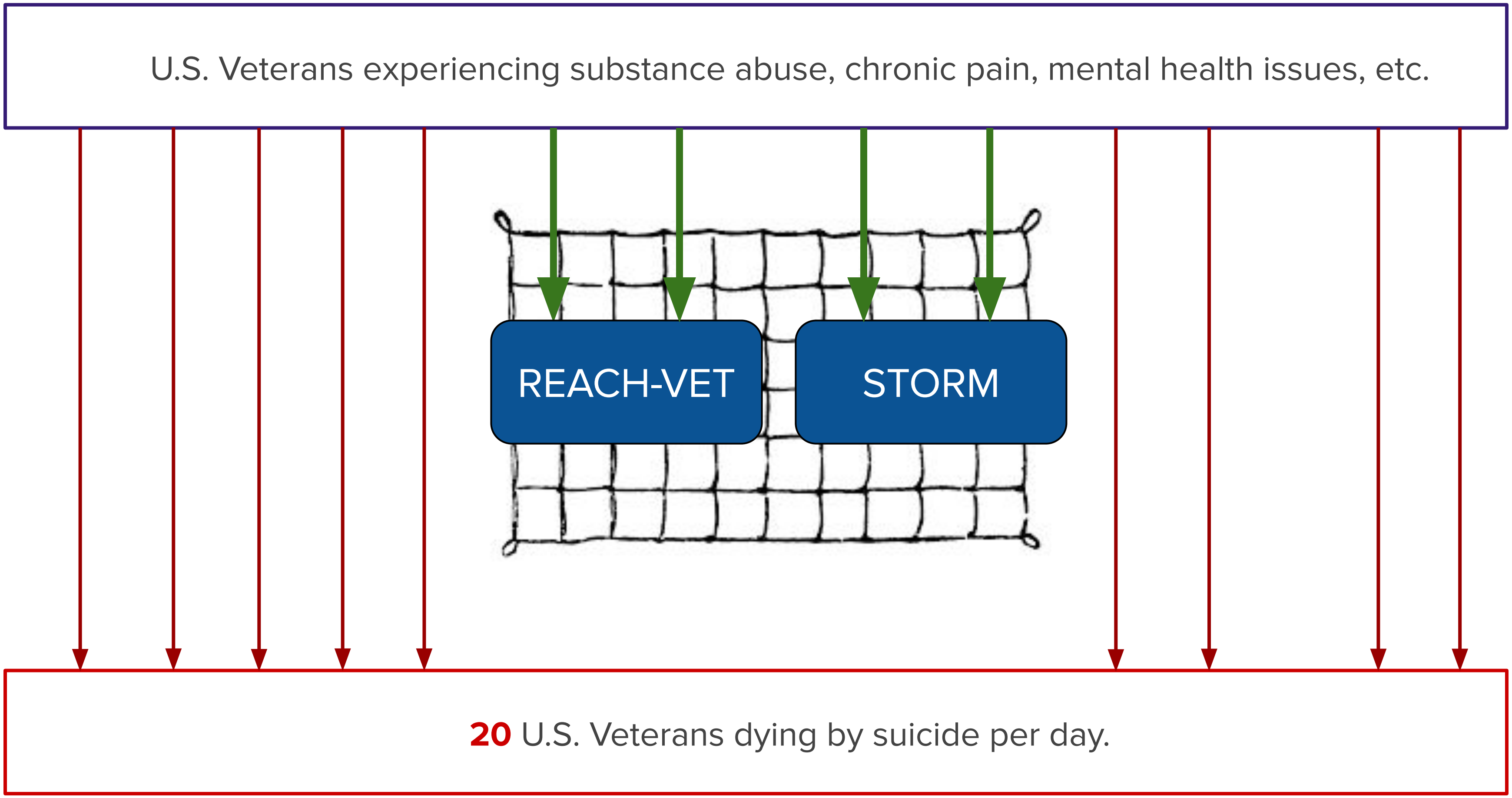
1. **Million Veterans Program**
    a. Current project and goals
2. **Related Work - Why Deep Learning?**
3. **Observational Medical Outcomes Partnership (OMOP) CDM**
4. **Data Preparation**
    a. Medical History Data Structure
    b. Medical Concept Embeddings
5. **Multi-Head Attention Longitudinal Model**
6. **Benchmarking on MIMIC-III**
    a. Prediction Tasks
    b. Results and Analysis

Million Veterans Program

- VA's Common Data Warehouse contains:
    - Clinical data for over **22** million patients.
    - **~3.4** billion unstructured medical documents.
    - Genomic data for **600,000** patients.

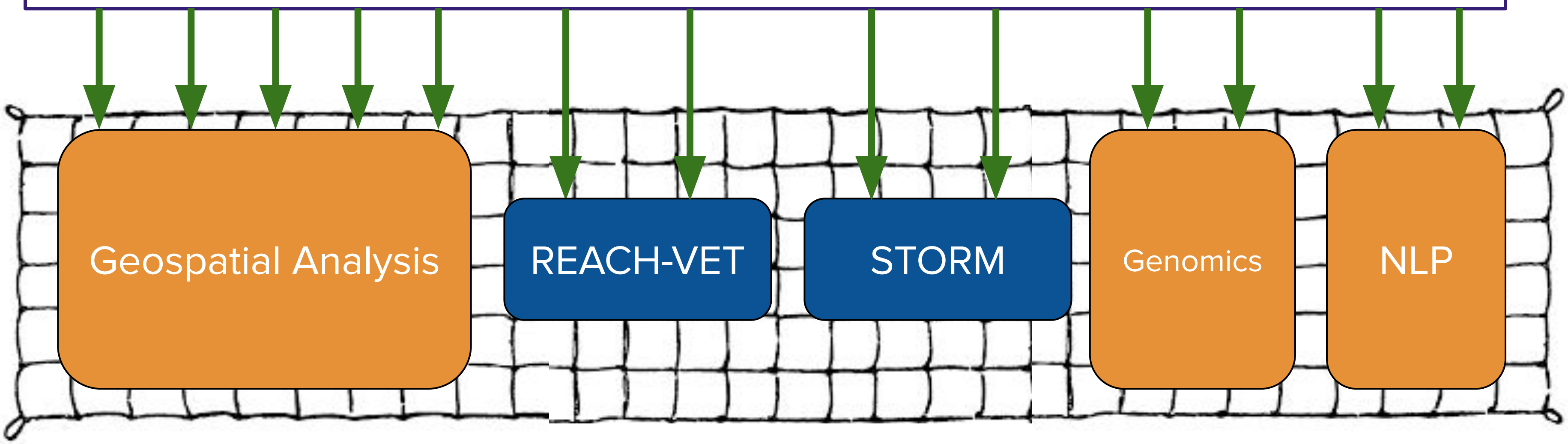
- 8 DOE Labs:
    - **Lawrence Berkeley**
    - Lawrence Livermore
    - Los Alamos
    - Sandia
    - Pacific Northwest
    - Oak Ridge
    - Brookhaven
    - Argonne

MVP@LBNL: Suicide Prevention



MVP@LBNL: Suicide Prevention

U.S. Veterans experiencing substance abuse, chronic pain, mental health issues, etc.



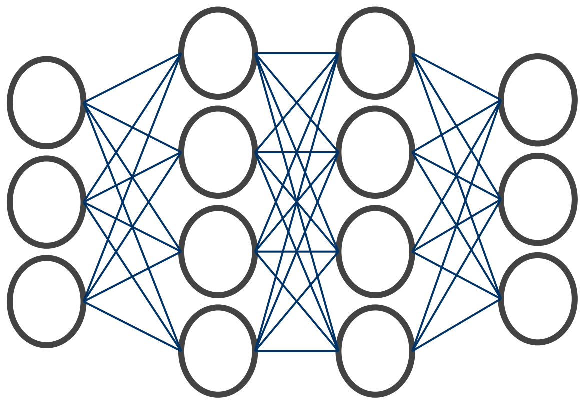
0 U.S. Veterans dying by suicide per day.

Deep Learning For Suicide Prevention

Integration of Multimodal Data

- Structured/Unstructured Medical Records
- Geospatial Data
- Genomic Profiles
- Medical Imaging

Find Patterns Overlooked By Experts



Reduce Time Spent Feature Engineering

- Predict Suicide Risk At Patient Level
- Construct Veteran Suicide Profile
- Provide Decision Support To VA Doctors

Related Work - Applying DL To EHR

- Rajkomar et al.[1] used Recurrent Neural Networks to predict for **unplanned readmission and mortality**.
  - **Common data model:**
    - Fast Healthcare Interoperability Resource (FHIR).[2]
  - Recurrent Neural Network (i.e LSTM).[3]
    - Computationally expensive to train.
    - Limited scope (last 48 hours)
  - Large hyper parameter search for embedding categorical variables [4]:
    - >201, 000 GPU hours used to train final model

1. *Rajkomar et al.. (2018). Scalable and accurate deep learning with electronic health records. NPJ Digital Medicine, 1(1), 18.*
2. *Walinjka & Woods (2018). FHIR tools for healthcare interoperability. Biomedical J. of Scientific and Technical Research.*
3. *Choi et al. (2016). Using recurrent neural network models for early detection of heart failure onset. J. of the Am. Med. Informatics Assoc..*
4. *Guo & Berkhahn (2016). Entity embeddings of categorical variables. arXiv preprint arXiv:1604.06737.*

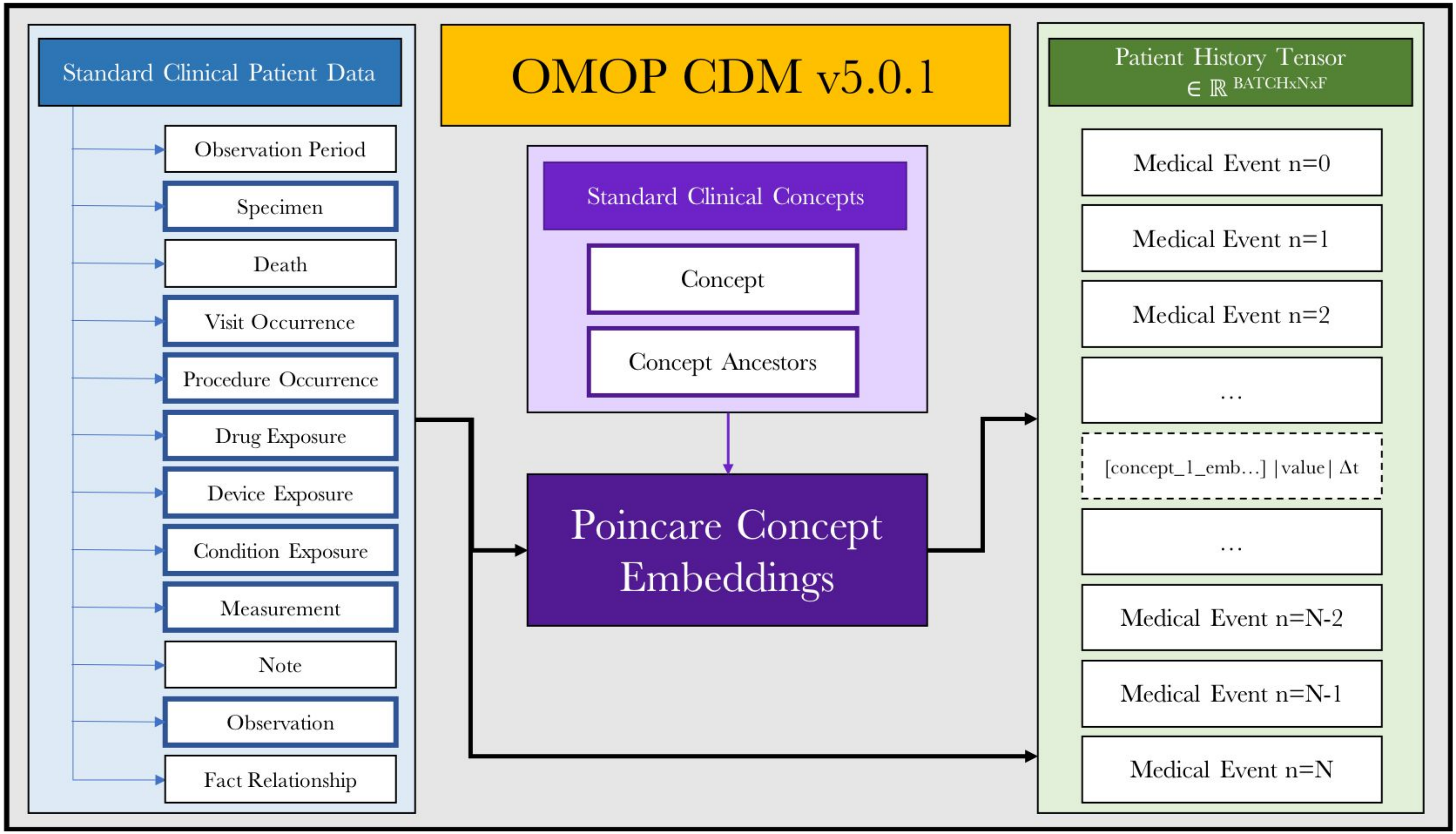
Can expanding the scope of the input representation to include common medical context and longer timelines improve a model's ability to predict medical outcomes?

OMOP Data Model

- The Observational Medical Outcomes Partnership (OMOP) CDM [5] :
    - Designed with data research as its primary purpose.
    - Addresses bottleneck in predictive medicine:
        - Lack of standardize data between medical facilities.
    - Clinical data are tagged with medical **concept identifiers**:
        - Harmonized between observations, diagnosis, medication, etc.
    - ***Schema currently in use by VA data science team.***

5. Hripcsak et al. (2015). Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. Studies in health technology and informatics, 216, 574.

Data Preparation

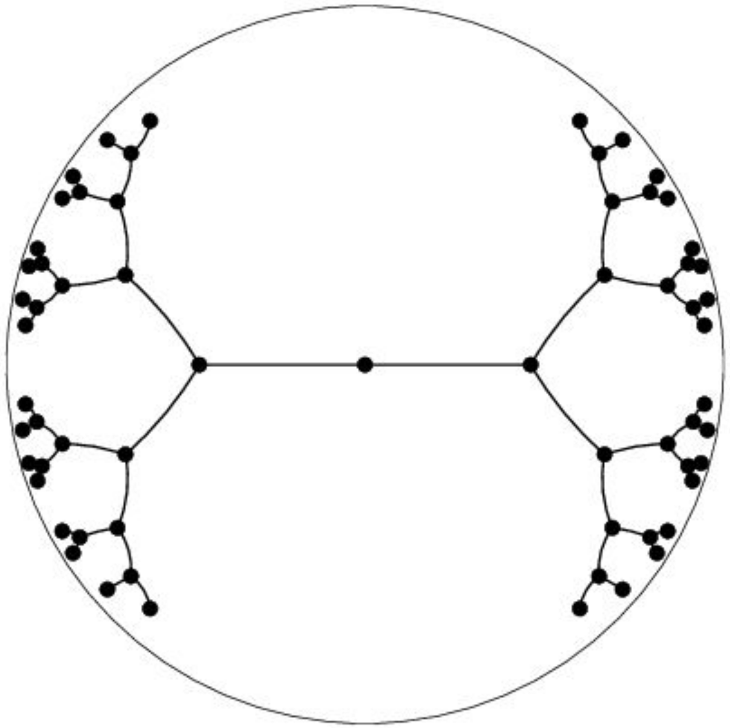


BERKELEY LAB

Medical Concept Embeddings

- OMOP's **Standardized Vocabularies** contains unique identifiers to medical concepts from source vocabularies included in the **Unified Medical Language System** (UMLS).
- OMOP's **concept_ancestor** table contains:
    - **68** million hierarchical relation for
    - **3.3** million standard concepts.
- Method: Poincare Space Embeddings [6]
- Model scaled to fill Tesla V100.

6. *Nickel et al. (2017). Poincaré embeddings for learning hierarchical representations. In Advances in neural information processing systems (pp. 6338-6347).'*

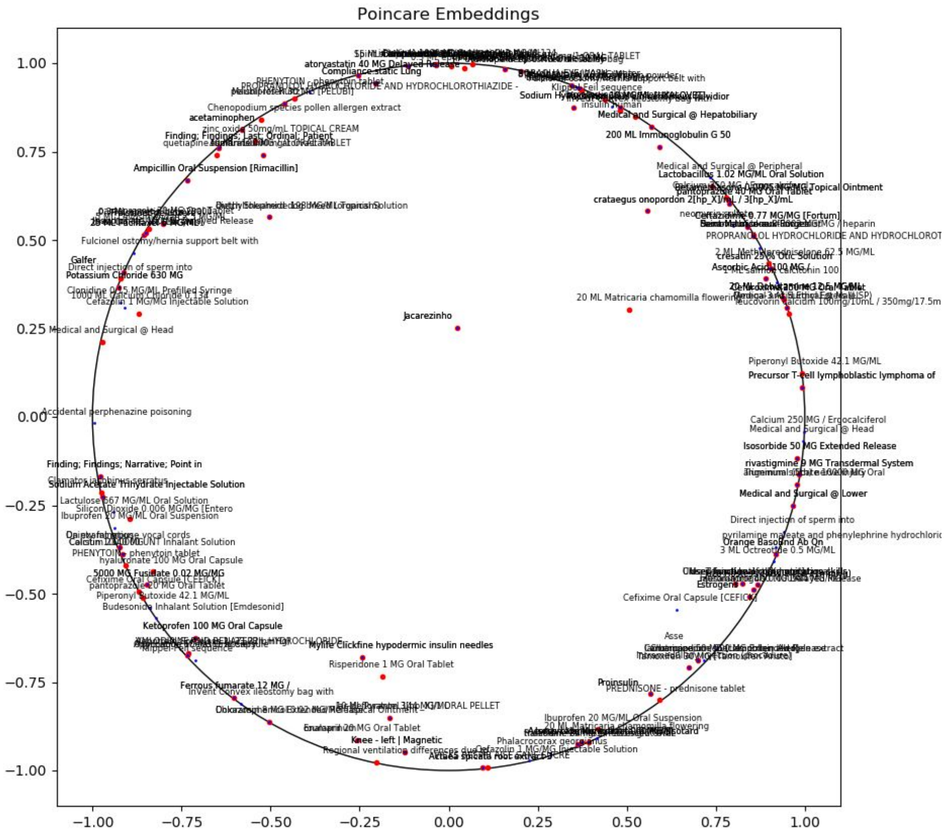


BERKELEY LAB

Medical Concept Embeddings

- OMOP's **Standardized Vocabularies** contains unique identifiers to medical concepts from source vocabularies included in the **Unified Medical Language System** (UMLS).
- OMOP's **concept_ancestor** table contains:
  - **68** million hierarchical relation for
  - **3.3** million standard concepts.
- Method: Poincare Space Embeddings [6]
- Model scaled to fill Tesla V100.

6. *Nickel et al. (2017). Poincaré embeddings for learning hierarchical representations. In Advances in neural information processing systems (pp. 6338-6347).'*



Visualization of Poincare embeddings for 500 randomly sampled OMOP concepts in 2-dimensions. Concepts with more common relations between other concepts appear closer to the origin.

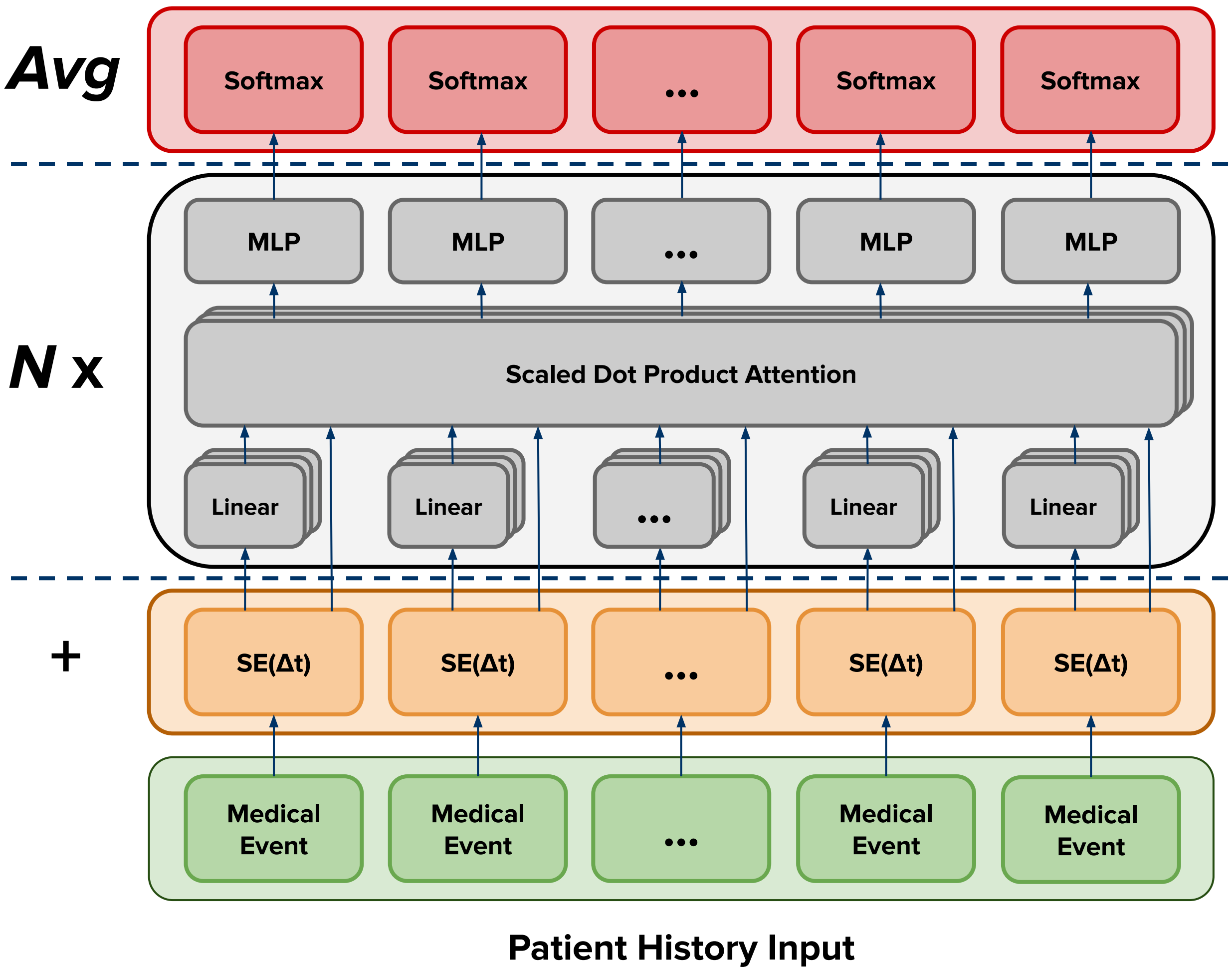
Multi-Head Attention For EHR

- "Transformers", the current state-of-the-art modelling method for natural language translation.
    - Attention is All You Need [7]
- **Multi-Head Attention** is used to learn intermediate representation of sequences.
- Networks learn to apply "attention" over sections of a sequence to predict the next event.
- Contemporary work applying attention to EHR:
    - Attend and Diagnose: Clinical Time Series Analysis Using Attention Models (2018) [8]

7. *Vaswani et al. (2017). Attention is all you need. In Advances in neural information processing systems(pp. 5998-6008).*
8. *Song et al. (2018, April). Attend and diagnose: Clinical time series analysis using attention models. In Thirty-Second AAAI Conference on Artificial Intelligence.*

- Benefits:
    - Reduces data sparsity (i.e. dense representation).
    - Process patient histories with **varying granularities** of event occurrences.
- Drawbacks:
    - Large number of medical events can be recorded for even just 1 medical admission.
    - Example: Average of ~ 5,000 events per admission in MIMIC-III.

Model Architecture



Class Probability Output:

Cross Entropy Loss

$$CE = -\sum_{i}^{C} t_i \log(f(s)_i)$$

Multi-Head Attention Encoder Stack:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

$$\mathbf{h}^l = f\left[\frac{\mathbf{g}}{\sigma^l} \odot (\mathbf{a}^l - \mu^l) + \mathbf{b}\right]$$

Sequential Encoding:

$$SE(\Delta t, 2i) = \sin(\Delta t / \boldsymbol{\lambda}^{2i/d})$$

$$SE(\Delta t, 2i+1) = \cos(\Delta t / \boldsymbol{\lambda}^{2i/d})$$

$$\boldsymbol{\lambda} = 700{,}800 / 2\pi$$

Hours in an 80 year lifespan.

Retrospective Study: MIMIC-III

- MIMIC-III [9] contains data of ICU stays of the Beth Israel Deaconess Medical Center from 2001-2012.
- Highly imbalanced dataset:
  - **608** suicide related patients; 45,913 non-suicide related patients
  - Not enough data for meaningful results, so we tackled the following tasks:

- Can we predict the likelihood that a patient will **die or have an unplanned readmission** after being discharged from the ICU?
- Can we predict the likelihood that a patient will **die in-hospital** using records from the first 48-hours after admission?
- Both tasks are still highly unbalanced.

9. *Johnson et al.. (2016). MIMIC-III, a freely accessible critical care database. Scientific data, 3, 160035.*

30-Day Unplanned Readmission Results

- Dataset split provided by Lin et al. [10] :
  - Trained for 50 epochs
  - Batch Size: 32; Learning Rate: 0.001 ------ 22.3 average batch updates per second

Tesla V100 GPU: ~ 2.5 million examples per hour

10. *Lin et al.(2019). Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. PloS one, 14(7), e0218942.*

Model	Feature Set	Dataset	min(Se, P+)	AUPRC	AUROC
Rajkomar et al.'s LSTM	Last 48 hours FHIR	UCSF Medical DB *[216,221 Admissions]*	-	-	0.75
Lin et al.'s LSTM+CNN	Last 48 hours: ICD9 + Demographics + expert selected features	**MIMIC-III**	0.367	0.4911	0.791
OMOP Multi-Head Attention	Sampled Full History: OMOP Schema	**MIMIC-III**	**0.5878**	**0.6304**	**0.8519**

In-Hospital Mortality Results

- Dataset split provided by Song et al. :
  - Trained for 100 epochs
  - Batch Size: 32; Learning Rate: 0.001 ------ 23.1 average batch updates per second

Model	Feature Set	Dataset	min(Se, P+)	AUPRC	AUROC
Rajkomar et al.'s LSTM	Last 48 hours: FHIR	UCSF Medical DB *[216,221 Admissions]*	-	-	0.95
Lin et al.'s LSTM+CNN	48 hours after Admission: ICD9 + Demographics + expert selected features	**MIMIC-III**	0.516	0.533	**0.851**
OMOP Multi-Head Attention	Sampled Full History up to 48 hours after Admission: OMOP Schema	**MIMIC-III**	**0.623**	**0.653**	0.837

Challenges and Ongoing Work

- **Access to MVP data has been established as of August 2019.**
  - Currently focusing on predicting risk of suicide.
- **On the model front:**
  - Scale deep learning to data spanning 20 years.
  - Capturing subtle concepts such as social isolation.
  - Integrate genomic data, medical notes, images.
- **On the results front:**
  - Physicians will want:
    - Uncertainty quantification
    - Interpretable methods: Workflow allows mappings to raw OMOP

Acknowledgements

This work was supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Thank You