BERTSurv: BERT-Based Survival Models for Predicting Outcomes of Trauma Patients

Yu Deng, Ph.D

Statistical Innovation Group, AbbVie Inc., North Chicago

Acknowledgement

- Yun Zhao, Qinghang Hong, Xinlu Zhang, Yuqing Wang, Linda Petzold (Department of Computer Science, University of California, Santa Barbara)
- Zhao, Yun, et al. "Bertsurv: Bert-based survival models for predicting outcomes of trauma patients." *arXiv preprint arXiv:2103.10928* (2021).

Disclaimer

- This publication was neither originated nor managed by AbbVie, and it does not communicate results of AbbVie-sponsored Scientific Research. Thus, it is not in scope of the AbbVie Publication Procedure (PUB-100) 10.

Outline

- Background
- BERTSurv
- Dataset
- Results
- Summary

Trauma Statistics

No.1 cause of death up to age of 45 in the USA

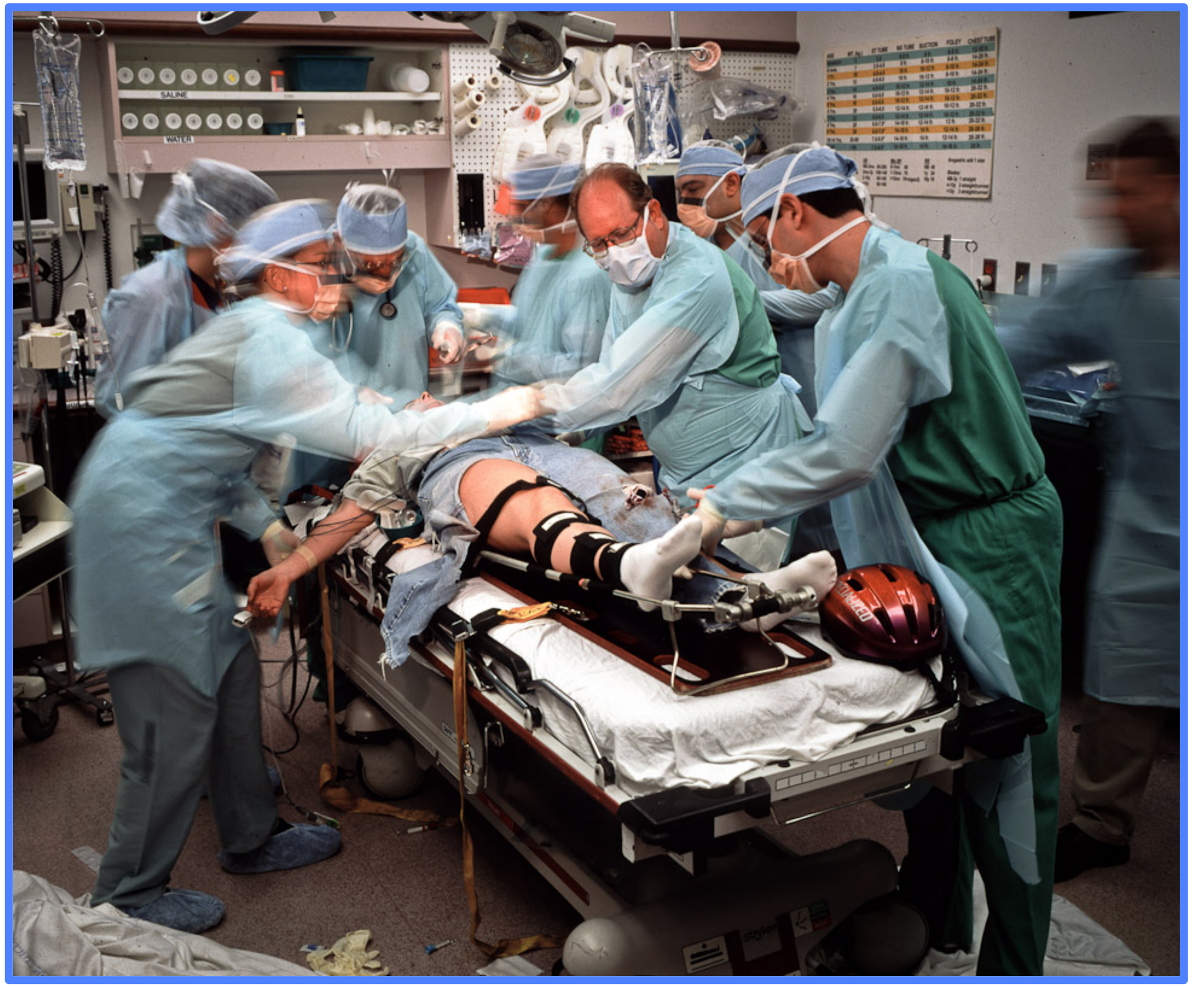
5 million death worldwide each year

No.1 cause of death from age 15 to 49 worldwide

41 million emergency department visits in the USA each year

Trauma

- Trauma deaths happen quickly
- Initial treatments and decision-making actions are required in first minutes or hours after injury
- The ICU has been found to be one of the sites where medical errors are most likely to occur
- Early and accurate prediction for trauma patient outcomes is essential for ICU decision making.



Data

Demographical data, history data

Measurements data

Lab tests data

Clinical notes

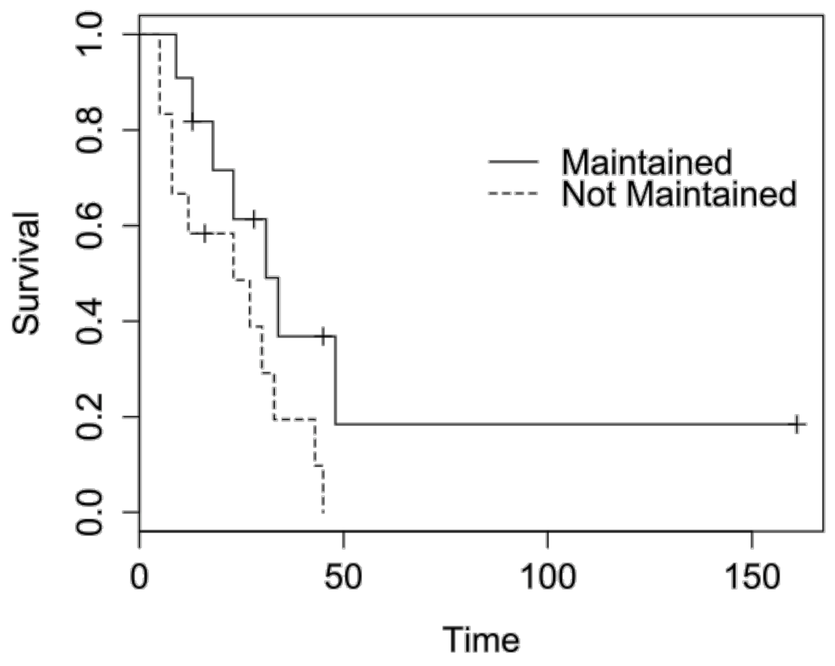
Survival Analysis

- Survival function: the probability of being alive just before t.

$$S(t) = P(T^* > t)$$

- Hazard function: the instantaneous rate of death at time t, given survival up to time t.

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < T^* \leq t + \Delta t | T^* > t)}{\Delta t} = \frac{f(t)}{S(t)}$$



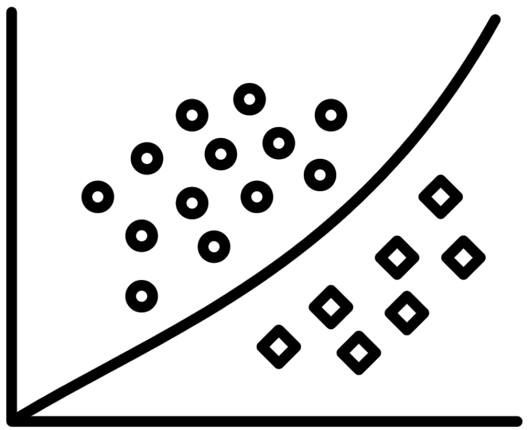
Mortality Prediction for Trauma Patients

Demographic Data, Disease History

Clinical Notes

Vital Signs

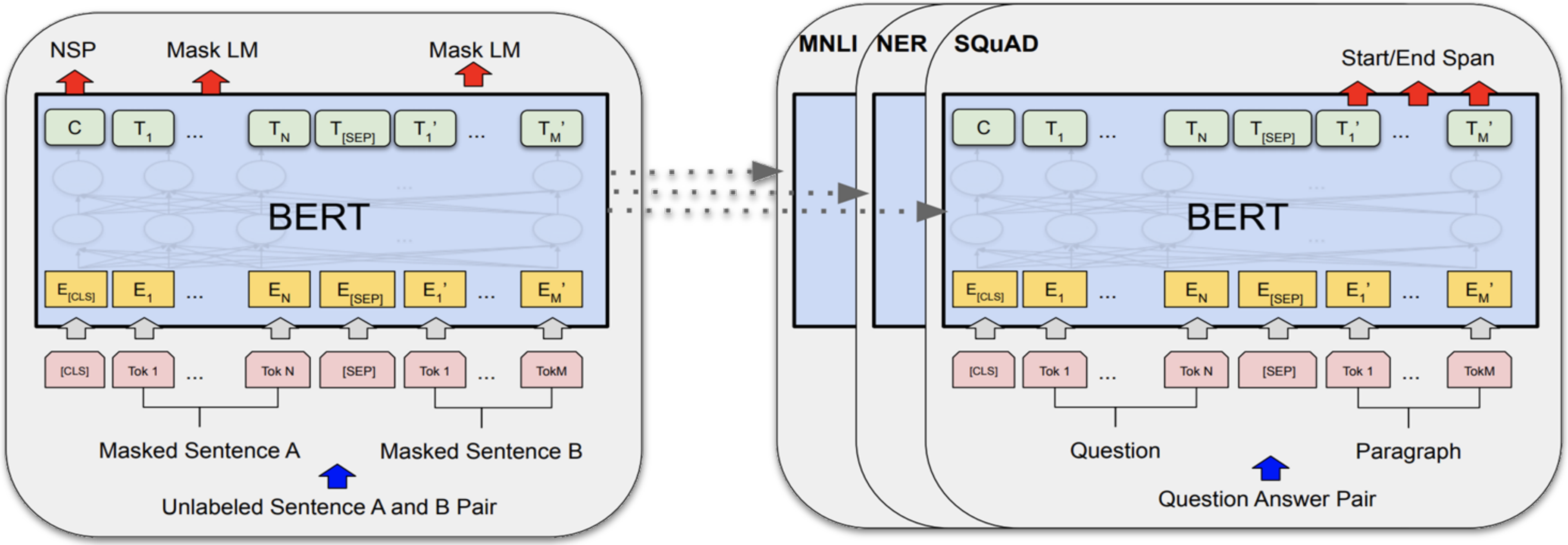
Lab Tests Data



Died in Hospital

Cured and Discharged from Hospital

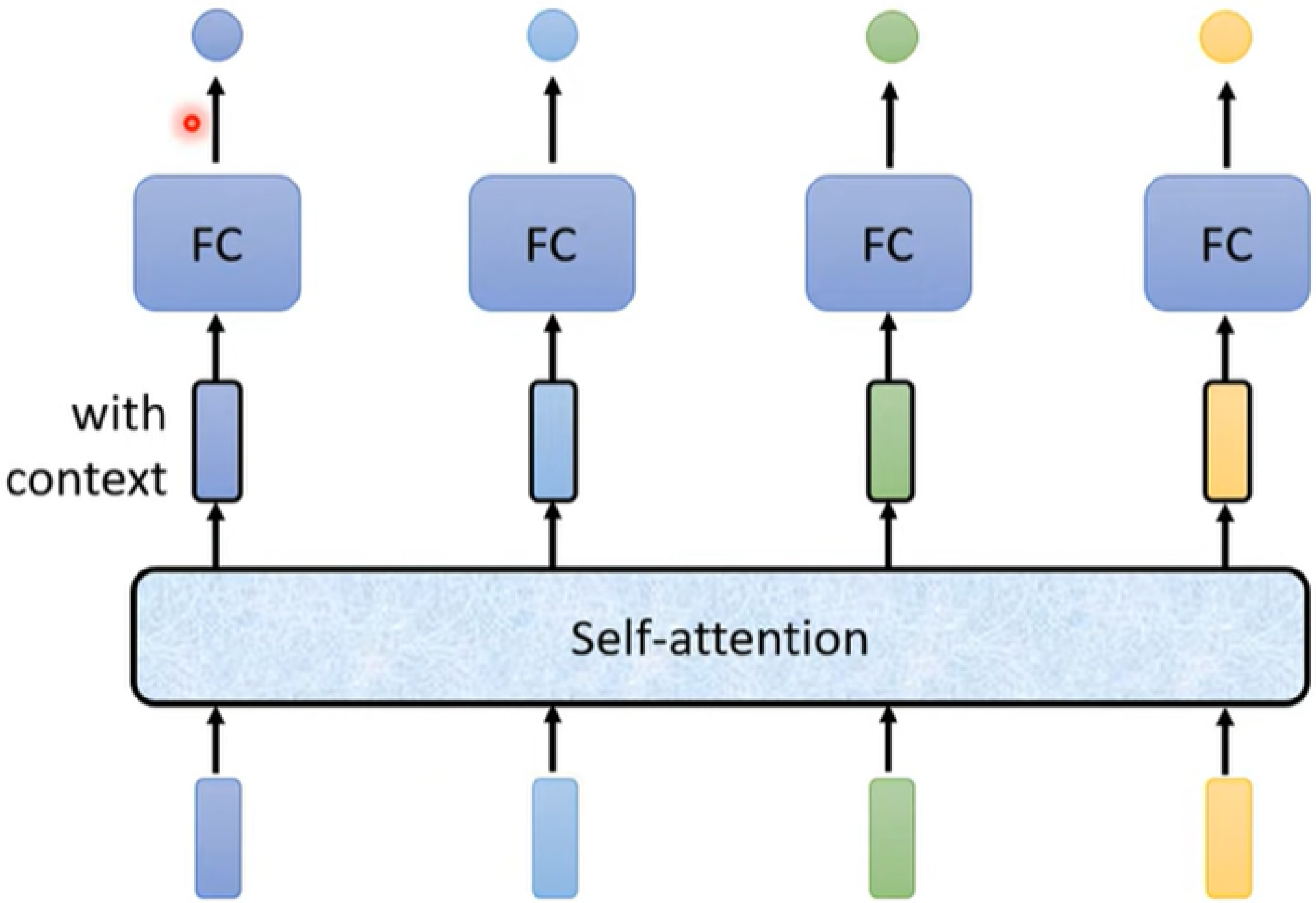
BERT(Bidirectional Encoder Representations from Transformers)



Pre-training Fine-Tuning

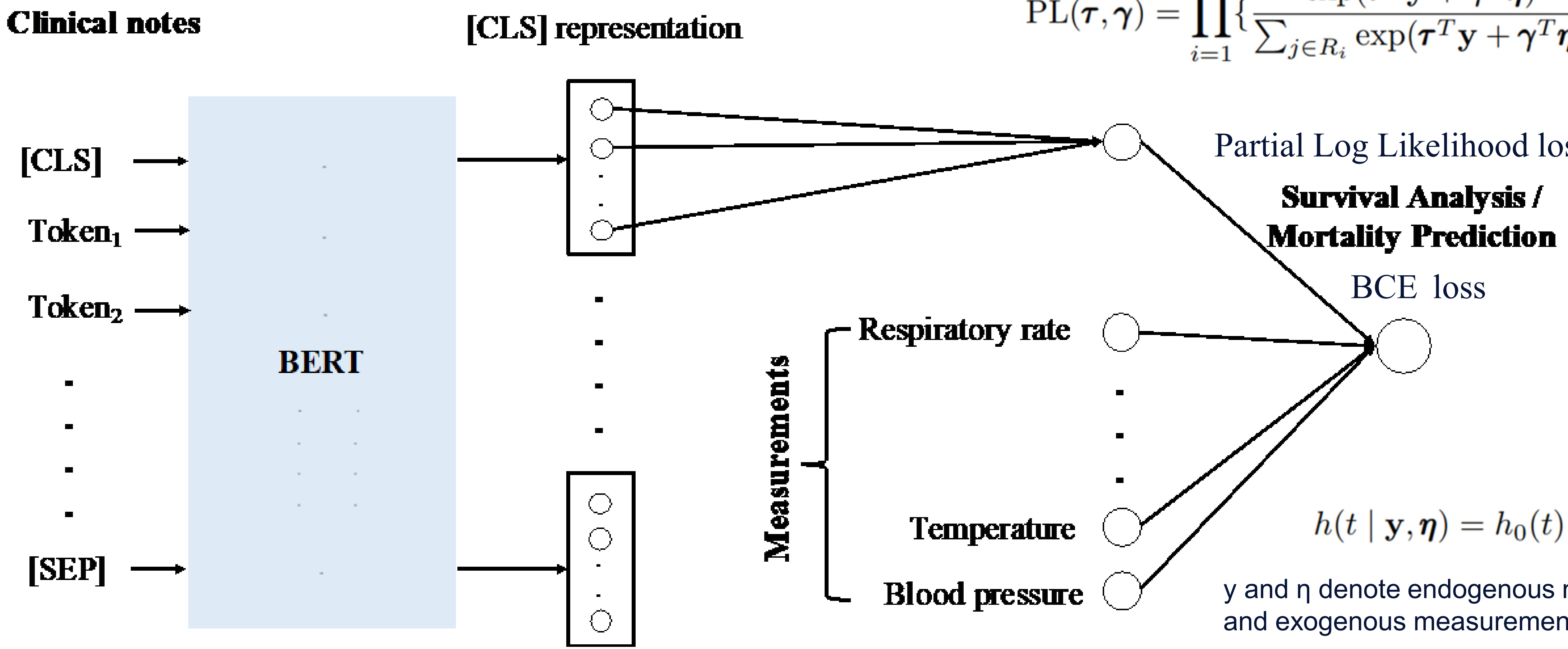
Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).
Liu, Yinhan, et al. "Roberta: A robustly optimized bert pretraining approach." arXiv preprint arXiv:1907.11692 (2019).
Conneau, Alexis, et al. "Unsupervised cross-lingual representation learning at scale." arXiv preprint arXiv:1911.02116 (2019).

Self-attention Mechanism



BERTSurv

$$\mathrm{PL}(\boldsymbol{\tau},\boldsymbol{\gamma}) = \prod_{i=1}^{n}\left\{\frac{\exp(\boldsymbol{\tau}^T\mathbf{y}+\boldsymbol{\gamma}^T\boldsymbol{\eta})}{\sum_{j\in R_i}\exp(\boldsymbol{\tau}^T\mathbf{y}+\boldsymbol{\gamma}^T\boldsymbol{\eta})}\right\}^{\Delta_i},$$



$$h(t \mid \mathbf{y},\boldsymbol{\eta}) = h_0(t)\exp(h_\theta(\mathbf{x}))$$

y and η denote endogenous measurements and exogenous measurements, respectively

$$\lambda_i(t) = \lambda_0(t)\exp\left(\beta_1 X_{i1} + \cdots + \beta_p X_{ip}\right) = \lambda_0(t)\exp(\boldsymbol{X}_i^T\boldsymbol{\beta})$$

Dataset

- MIMIC III dataset
  - Trauma patients are selected using the ICD-9 code (1860 ICU patients)
  - Measurements, demographic data, clinical notes, death outcome and time to death
  - sample class ratio between class 0 (discharge) and class 1 (death) is 1206 : 654.
- Preprocessing

Aggregate the measurements over the first 4 hours of an admission

→ Select 21 most common measurement features

→ Select 4 demographic features (weight, gender, ethnicity and age)

→ Remove missing value proportion > 0.4, MICE imputation

- blood pressure, temperature, respiratory rate, arterial PaO2, hematocrit, WBC, creatinine, chloride, lactic acid, BUN, sodium (Na), glucose, PaCO2, pH, GCS, heart rate, FiO2, potassium, calcium, PTT and INR (65% overlap with APACHE III score)

Notes Preprocessing

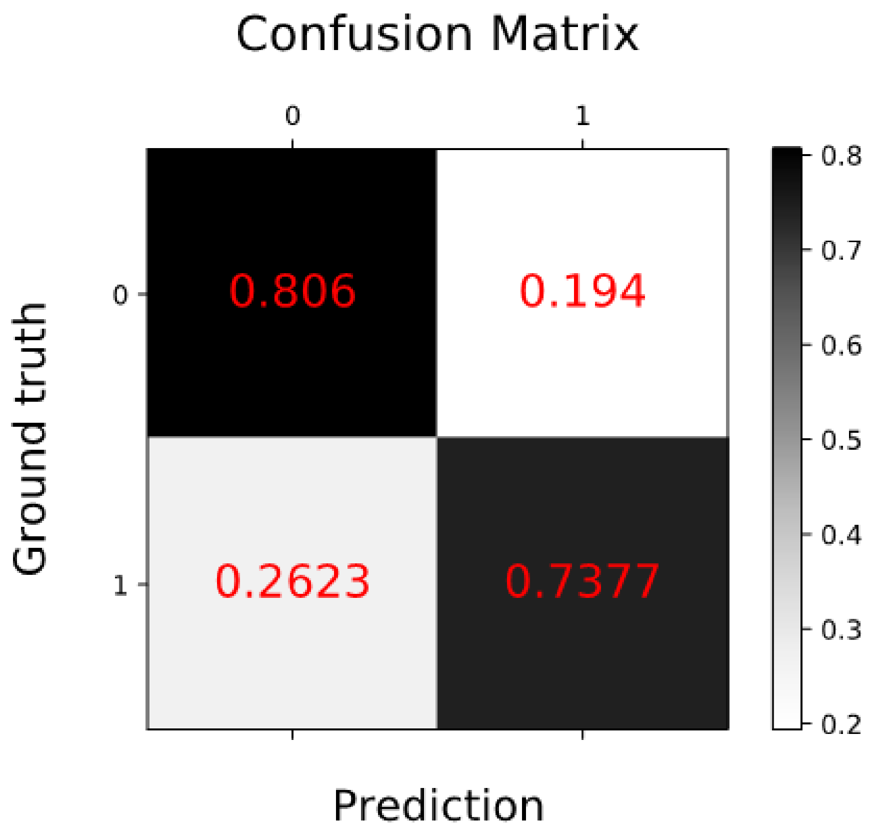
Concatenate the clinical notes over the first 4 hours

Remove formatting, numbers and non-punctuation symbols

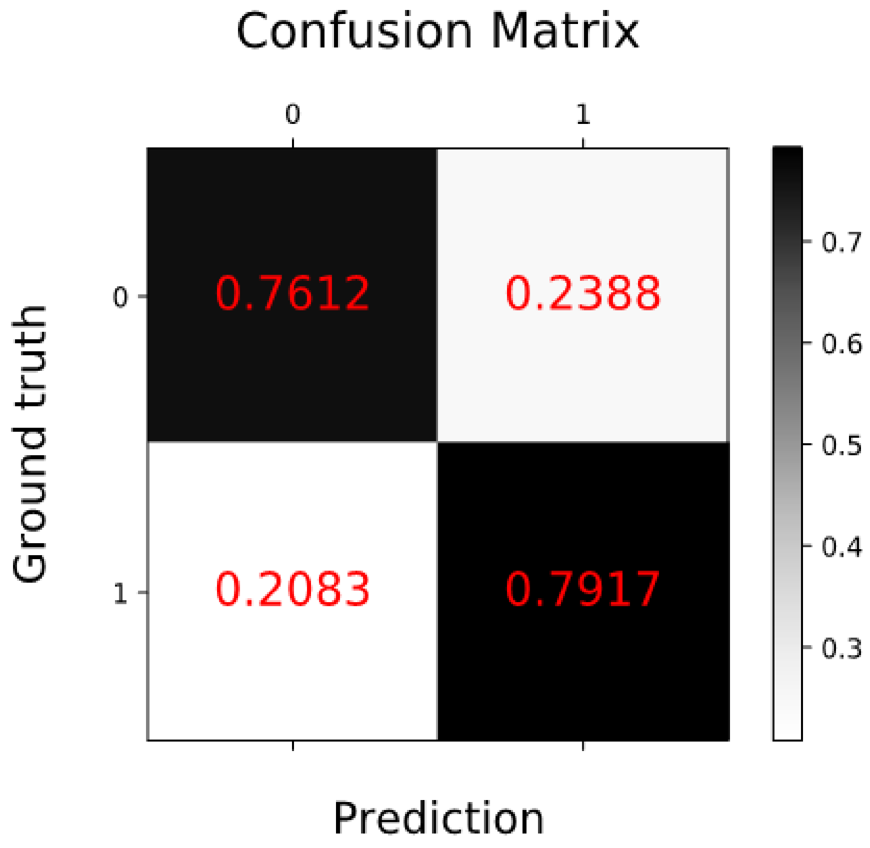
Remove stopwords (e.g. a, the, above)

Remove punctuations, morphological affixes

Mortality Binary Classification (BCE Loss) Confusion Matrix

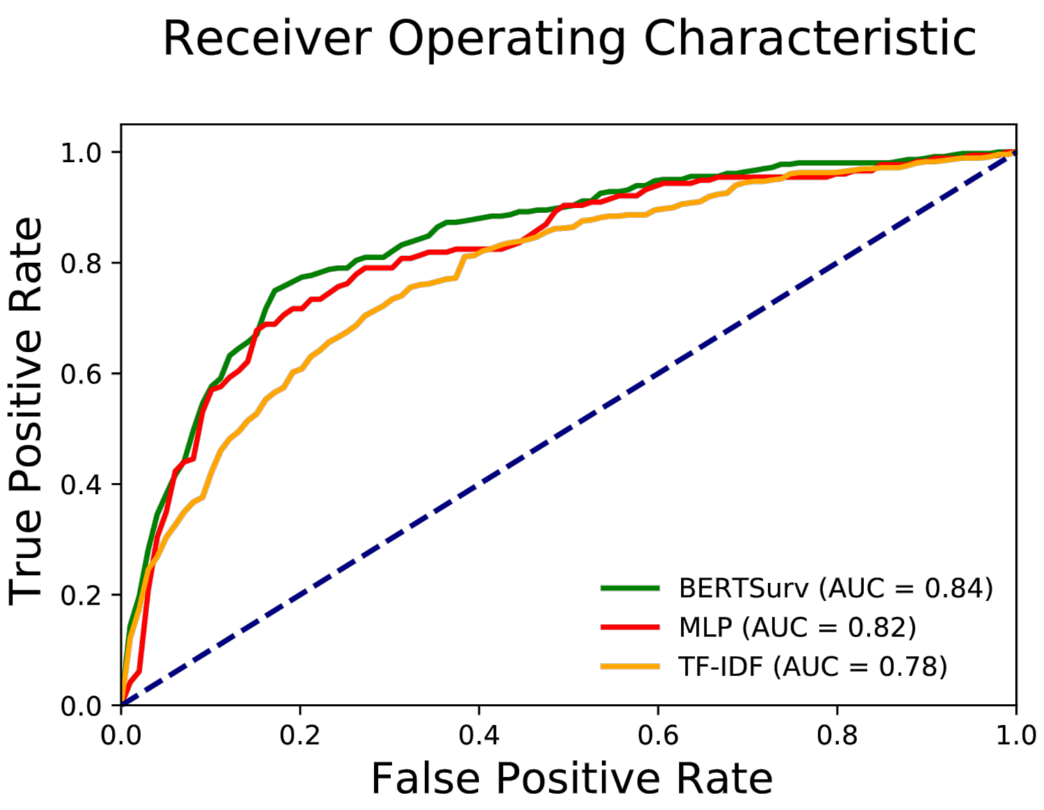


5 Fold Cross Validation

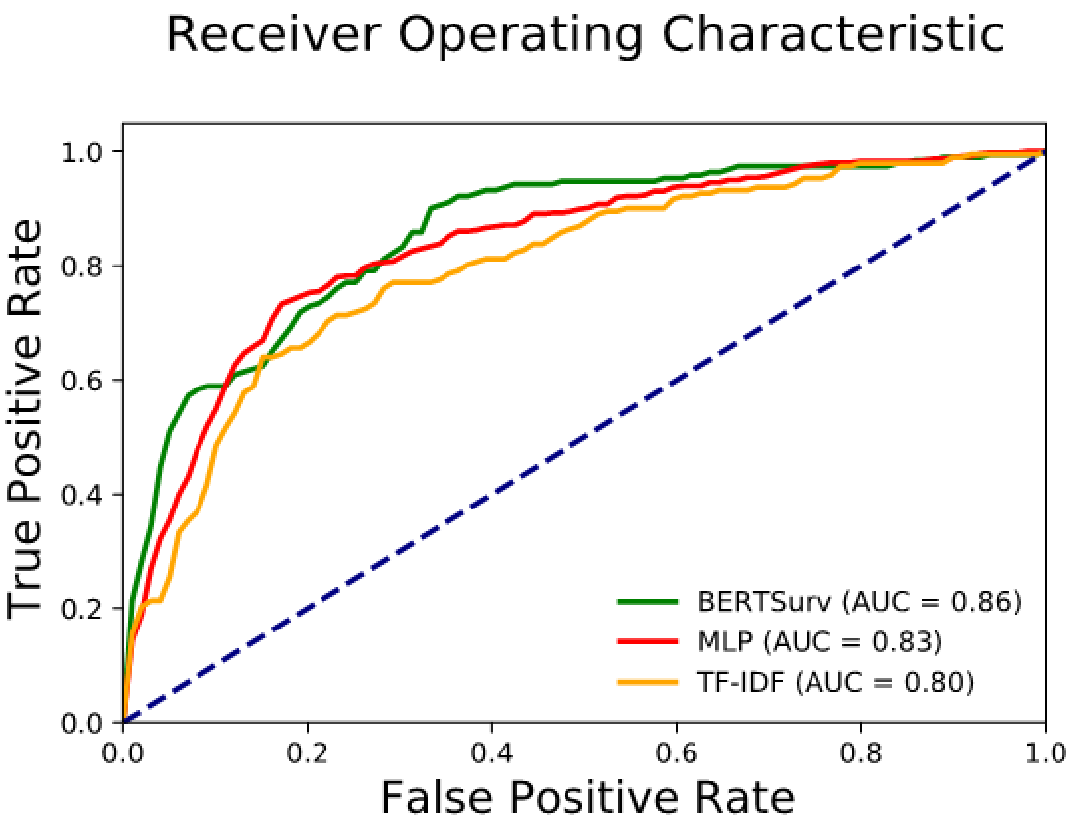


Test

Mortality Binary Classification (BCE Loss) ROC



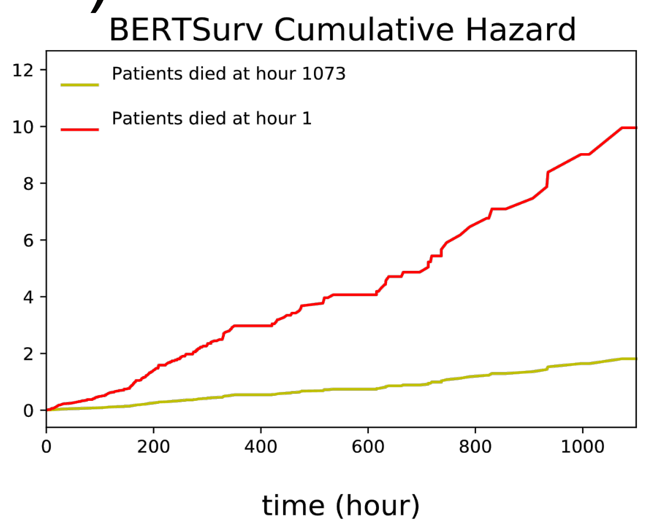
5 Fold Cross Validation

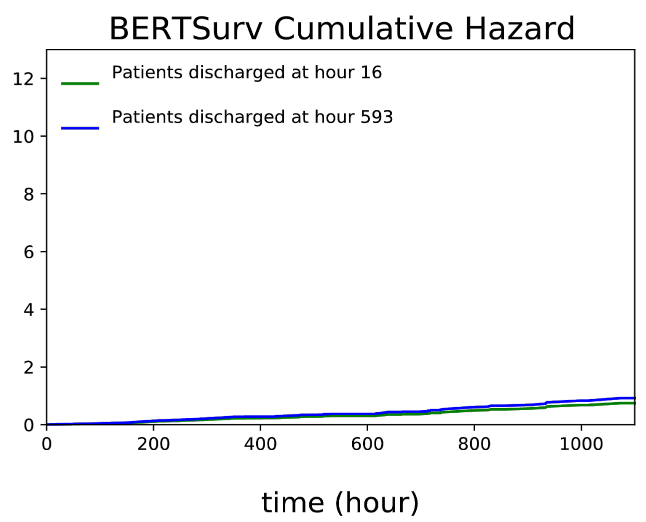


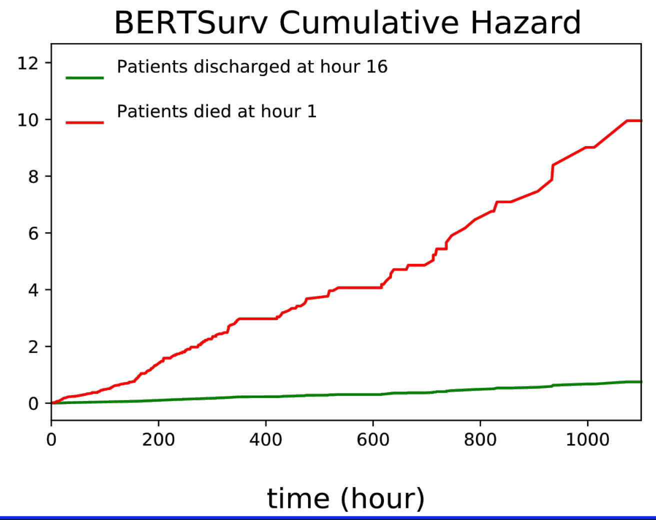
Test

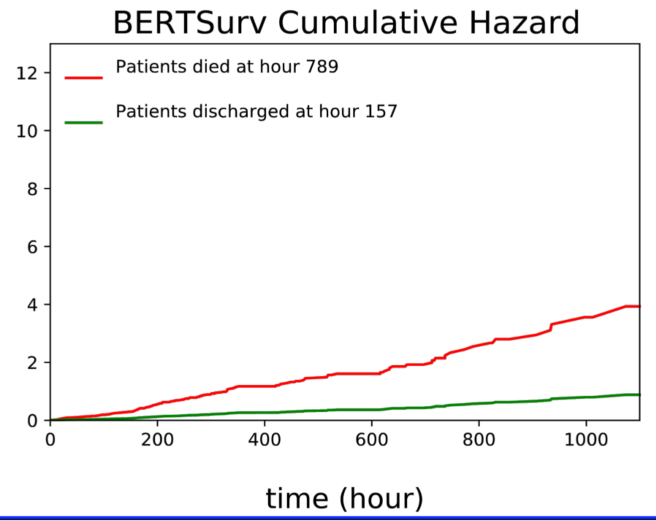
Survival Predictions (PLL loss)

- BERTSurv: C-index = 0.7
- Cox model: C-index = 0.68







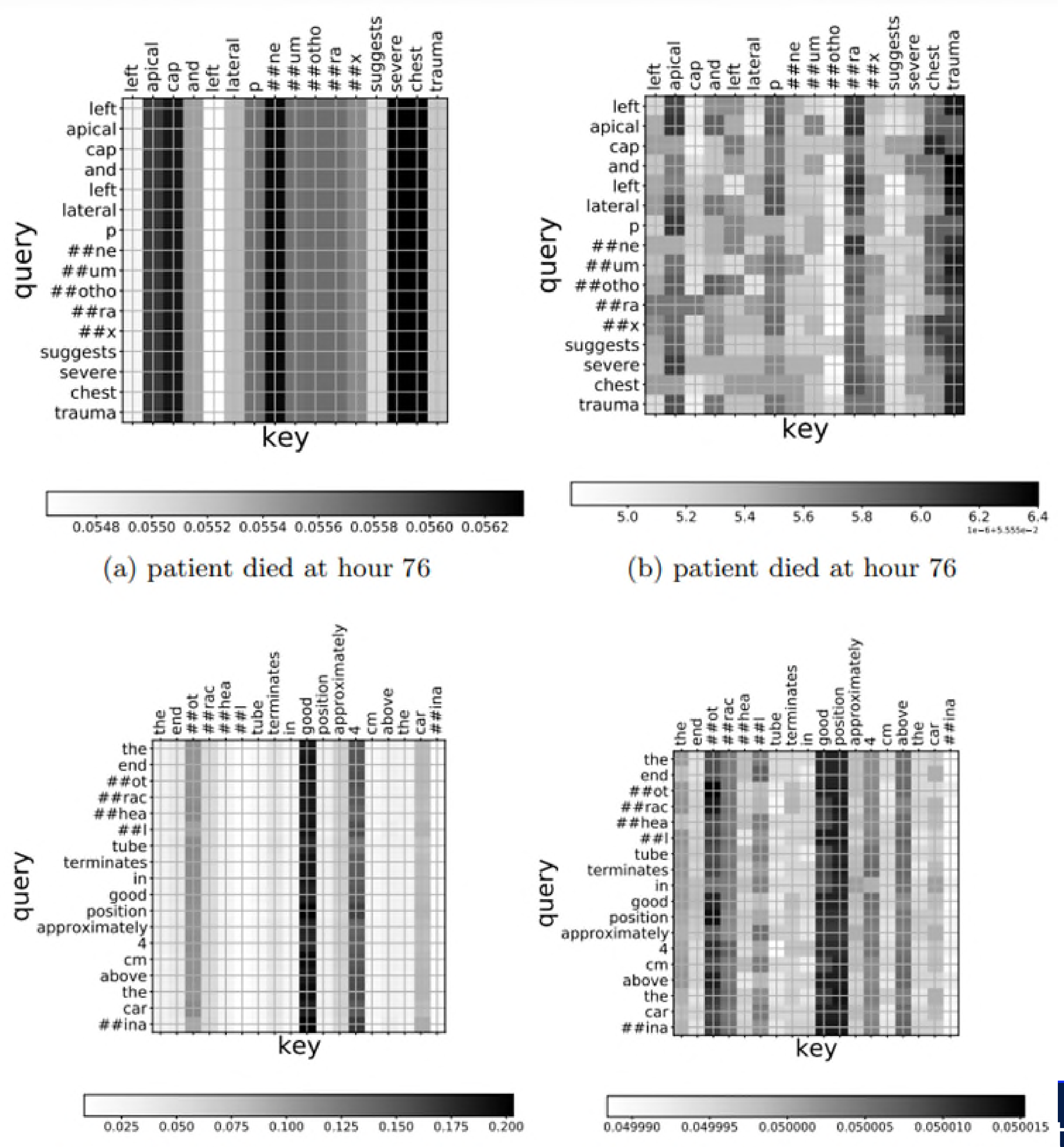


BERT Visualization

“the endotracheal tube terminates in **good** position approximately 4 cm above the carina.”

discharged at hour 85

“left apical cap and left lateral pneumothorax suggests **severe chest** trauma .”



(a) patient died at hour 76

(b) patient died at hour 76

(c) patient discharged at hour 85

(d) patient discharged at hour 85

died at hour 76

Summary

- We propose BERTSurv: a BERT-based deep learning framework to predict the risk of death for trauma patients.
- We evaluate BERTSurv on the trauma patients in MIMIC III. BERTSurv achieved a C-index of **0.7** on trauma patients, which outperforms a Cox model with a C-index of 0.68.
- We extracted patterns in the clinical texts with attention mechanism visualization and correlated the assigned weights with survival outcomes.

References

- Goff, David C., et al. "2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines." *Journal of the American College of Cardiology* 63.25 Part B (2014): 2935-2959.
- Katzman, Jared L., et al. "DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network." *BMC medical research methodology* 18.1 (2018): 1-12.
- Gensheimer, Michael F., and Balasubramanian Narasimhan. "A scalable discrete-time survival model for neural networks." *PeerJ* 7 (2019): e6257.
- Ching, Travers, Xun Zhu, and Lana X. Garmire. "Cox-nnet: an artificial neural network method for prognosis prediction of high-throughput omics data." *PLoS computational biology* 14.4 (2018): e1006076.
- Harrell, Frank E., et al. "Evaluating the yield of medical tests." *Jama* 247.18 (1982): 2543-2546.
- Zeng, Zexian, et al. "Natural language processing for EHR-based computational phenotyping." *IEEE/ACM transactions on computational biology and bioinformatics* 16.1 (2018): 139-153.
- Chibnik, Lori B., E. M. Massarotti, and Karen H. Costenbader. "Identification and validation of lupus nephritis cases using administrative data." *Lupus* 19.6 (2010): 741-743.
- Li, Tingting, et al. "Development and validation of lupus nephritis case definitions using United States veterans affairs electronic health records." *Lupus* 30.3 (2021): 518-526.
- Aronson, Alan R. "Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program." *Proceedings of the AMIA Symposium*. American Medical Informatics Association, 2001.

Thank you!