Early Prediction of Sepsis using Time Series Forecasting

Jinghua Xu*, Natalia Minakova*, Pablo Ortega Sanchez*, Stefan Riezler

Department of Computational Linguistics, Heidelberg University

Presenter: Jinghua Xu
Corresponding: xu@cl.uni-heidelberg.de

***Equal contribution**

- Sepsis is a serious complication of an infection
- It is a leading cause of death the Intensive Care Units (ICU)
- Early detection of sepsis is crucial for patient survival [2]
- In this work, we:
  - Implement a rule-based **sepsis check** based on a widely accepted guideline for sepsis recognition (Sepsis-3) [3]
  - Use and refine an existing **transformer-based deep learning model [1]** for
    - **time series forecasting** to support the sepsis-check
    - **24-hour sepsis prediction** in fully data-driven setup

Sepsis-3 Check

Suspected Infection

- One of the key indicators for septic patients: a **suspected infection** [3] [19]
- A suspected infection is identified by orders for
  - blood cultures
  - antibiotics

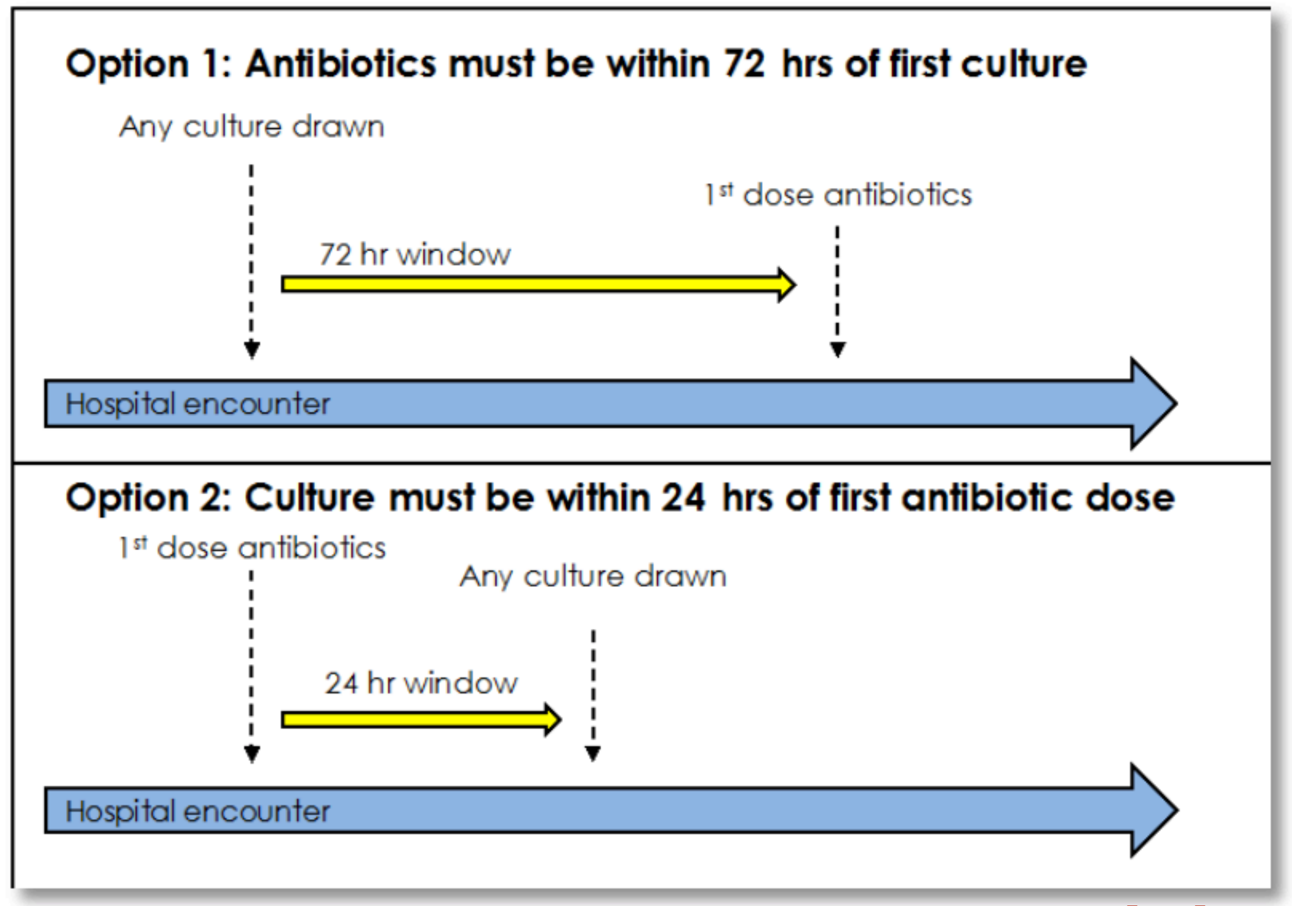

Figure source: supplementary materials of [19].

- Suspicion windows: [24 hours, 72 hours] [3] [19]
  - blood cultures taken first -> antibiotics need to be administered within 72 hours
  - antibiotics administered first -> blood cultures need to be taken within 24 hours

Sepsis-3 Check

Life-threatening Organ Dysfunction

- Another key indicator for sepsis: **life-threatening organ dysfunction**
- [3] suggests the Sequential [Sepsis-related] Organ Function Assessment (SOFA) [9]:
  - considers a variety of clinical and laboratory variables
- Based on [3]
  - SOFA score should be computed per hour:
    - Time (SOFA) = Time (SOFA score >= 2)
    - (considering the initial value to be zero if no organ dysfunction is known beforehand)

Sepsis-3 Check

Life-threatening Organ Dysfunction

- To identify Sepsis [3]
    - Time (SOFA) at most 48 hours before Time (Suspected Infection)

  OR

    - Time (Suspected Infection) 24 hours after Time (SOFA)
- Sepsis Window of [**x**, **y**]:
    - Time (SOFA) no more than **x** hours before

  OR

    - **y** hours after Time (Suspected Infection)

Sepsis-3 Check
SOFA

- In order to compute SOFA score, Glasgow Coma Scale need to be based on its components and the **mean arterial pressure**:

$$DBP + \frac{SBP - DBP}{3} \tag{2}$$

  - diastolic blood pressure (DBP)
  - systolic blood pressure (SBP)

Sepsis-3 Check

Implementation

TABLE X: Suspected infection components and corresponding variable names

component	variable name
time of blood cultures	Blood Cultures
time of antibiotics	Antibiotics

TABLE XI: SOFA components and corresponding variable names

component		variable name
Nervous System	Glasgow Coma Scale	GCS_eye, GCS_verbal, GCS_motor
Cardiovascular	Mean Arterial Pressure	DBP, SBP
	Administration of Vasopressors	Dopamine, Dobutamine, Epinephrine, Norepinephrine
Respiratory System	FiO2 [kPa]	FiO2
	Mechanical Ventilation	Mechanical ventilation
Coagulation	Platelet Count [x10^3/mu l]	Platelet Count
Liver	Bilirubin [mg/dl]	Bilirubin (Total)
Renal	Creatinine [mg/dl]	Creatinine Urine
	Urine [ml/day]	Urine

The absence of either the antibiotics or blood culture feature will always result in a negative sepsis label.

Data

Medical Information Mart for Intensive Care 3 (MIMIC-III) [14]

- MIMIC-III: a large dataset of de-identified health records of ICU patients
- 133 physiological features (Table IV in Appendix A)
- 2 static features: age & gender
- 1,407,430 clinical notes

TABLE II: String length and token counts in clinical notes included in our data.

	Avg.	Max.	Min.
String length	1673	55728	3
Num. tokens	316	11336	0

Data

Medical Information Mart for Intensive Care 3 (MIMIC-III) [14]

- Annotate sepsis label according to 23 sepsis-related ICD-9 codes (Table VII in Appendix B)
- 5288 septic patients (9.2%) and 51994 non-septic patients
- train/validation/test by 64: 16: 20 at patient level

TABLE I: Number of septic/non-septic patients/ICU stays in train/validation/test data.

Data	Non-septic patients	Septic patients	Non-septic ICU stays	Septic ICU stays
Train	26452	2124	33191	3360
Valid	6594	551	8358	904
Test	8296	635	10445	1024

Model

Self-supervised Transformer for Time-Series (STraTS) [1]

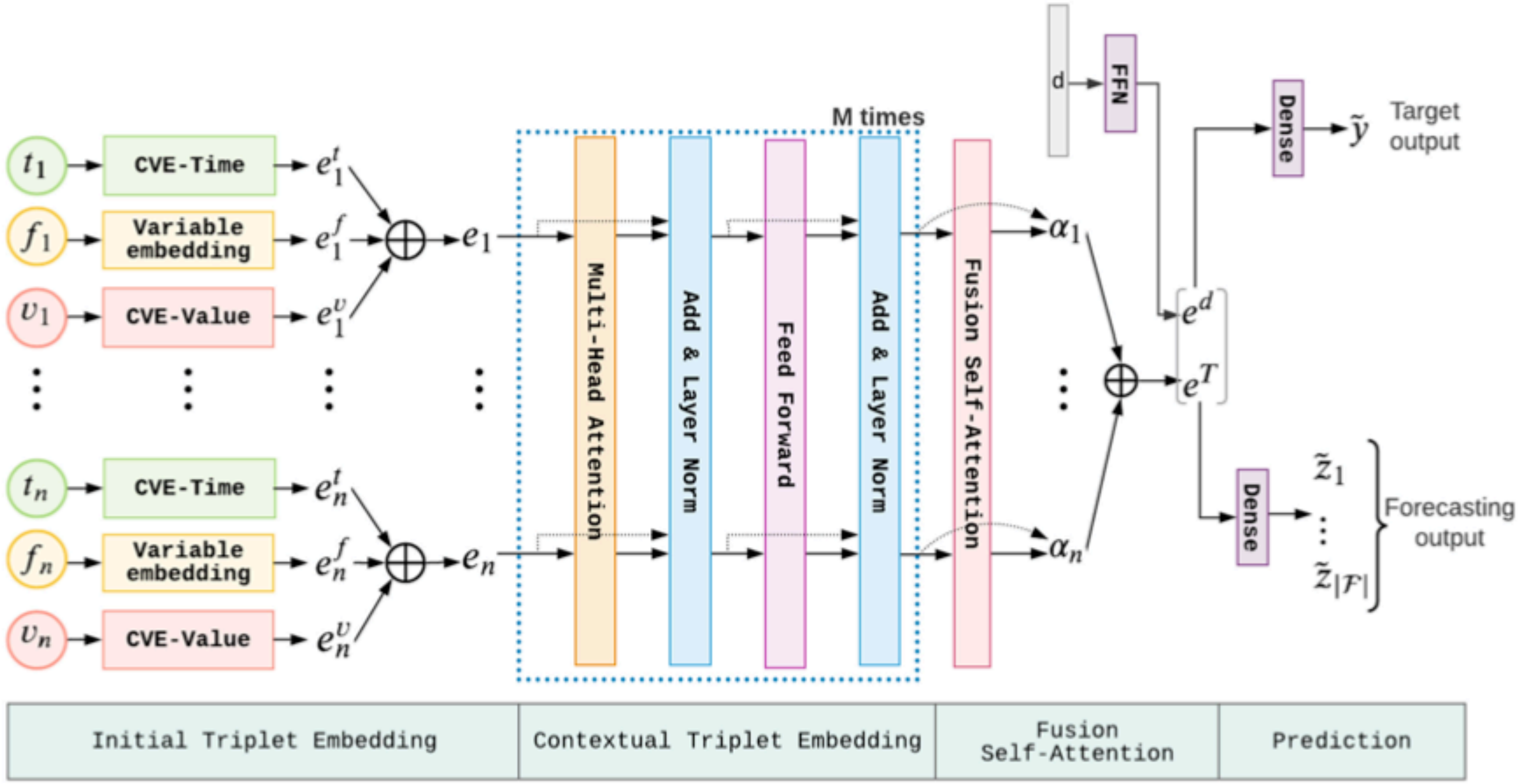

Fig. 3. The overall architecture of the proposed STraTS model. The Input Triplet Embedding module embeds each observation triplet, the Contextual Triplet Embedding module encodes contextual information for the triplets, the Fusion Self-Attention module computes time-series embedding, which is concatenated with demographics embedding and passed through a dense layer to generate predictions for target and self-supervision (forecasting) tasks.

Figure 3 in [1].

Our Refined Model

Self-supervised Transformer for Time-Series (STraTS) [1] + ClinicalBERT [5]

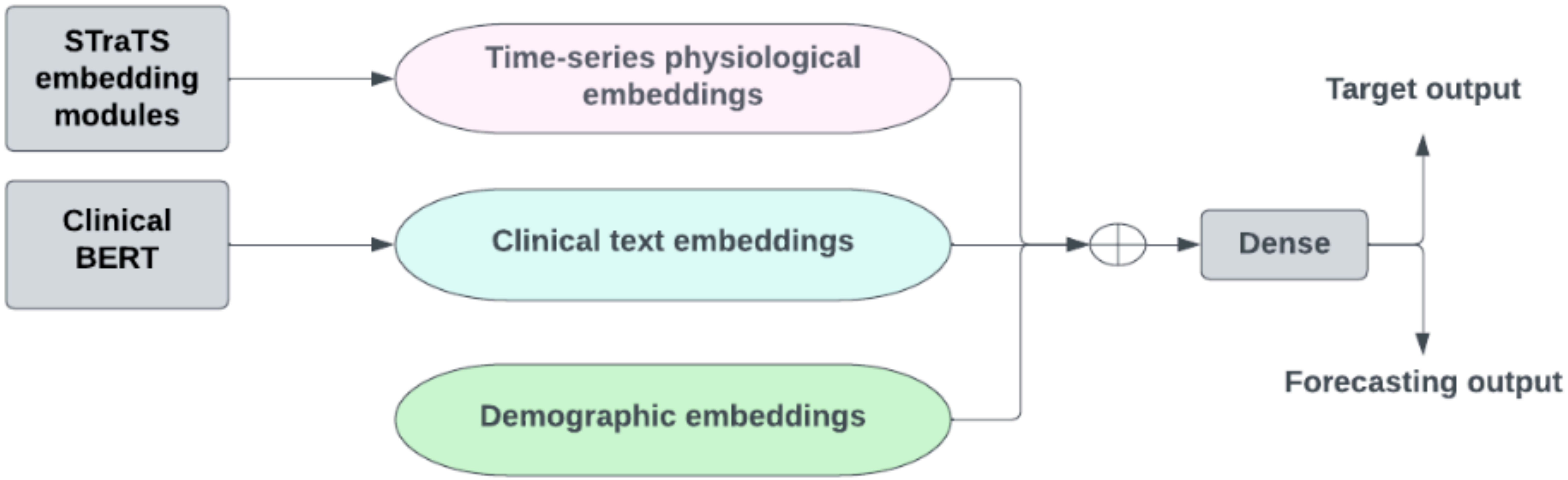

Fig. 1: STraTS + Clinical Text Embedding Architecture.

Results

STraTS Forecasting

2-hour forecasting window

TABLE IV: Masked MSE (mean squared error) on test and validation data for STraTS and STraTS + Text models.

Model	Test	Validation
STraTS	5.2455	5.2048
STraTS + Text	5.2493	5.5922

STraTS Classification
24-hour Sepsis Prediction

TABLE III: Sepsis prediction performance on MIMIC-III dataset. The results show mean and standard deviation of the metrics after repeating the experiment 10 times by sampling 50% labeled data each time.

Model	ROC-AUC	PR-AUC	min(Re,Pr)
STraTS	0.891 ± 0.003	0.500 ± 0.009	0.507 ± 0.100
STraTS + Text	0.889 ± 0.002	0.491 ± 0.008	0.492 ± 0.008

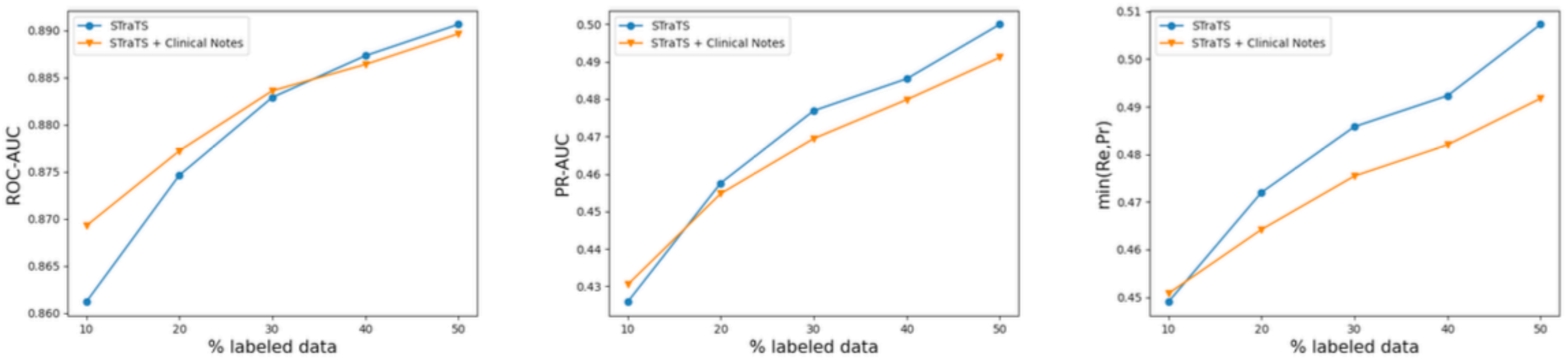

Fig. 2: Sepsis prediction performance on MIMIC-III dataset for different percentages of labeled data averaged over 10 runs.

Sepsis Check

TABLE V: 1: experiments were conducted with a suspicion window of 48 and 72 hours, and a sepsis window of 24 and 12 hours. 2: experiments were conducted with a suspicion window of 24 and 96 hours, and a sepsis window of 24 and 12 hours.

Experiment	F1 score[1]	Sensitivity	Specificity
1.1 observed-only	**0.874505**	0.262004	0.951701
1.2 observed+forecast	0.873794	**0.298538**	0.942567
2.1 observed-only	0.87309	0.244258	0.953528
2.2 observed+forecast	**0.873677**	**0.276617**	0.947134

1 *We report weighted average F1 score.*

Conclusion

- **Rule-based sepsis check**
  - Good by itself
  - Time series forecasting improved **sensitivity**
    - Ongoing work is extending the forecasting window
  - Sparseness of patient data greatly impacts potential performance
- **Data-driven model:**
  - Good by itself for both **24-hour sepsis prediction & time series forecasting**
  - Newly added text embedding module:
    - Augmented data in case of insufficient amount of labelled data
    - Ongoing work is seeking to improve the text embedding module:
      - Alternative LMs for text embedding
        - RoBERTa-large-PM-M3-Voc (Pre-trained on PubMed and PMC and MIMIC-III with a BPE Vocab learnt from PubMed) [20]
        - Mimic-longformer (Longformer trained on MIMIC-III clinical notes) [21]
      - Architecture-wise changes to the model

References

[1] S. Tipirneni and C. K. Reddy, “Self-supervised transformer for sparse and irregularly sampled multivariate clinical time-series,” 2021. [Online]. Available: https://arxiv.org/abs/2107.14293

[2] K. E. Rudd, S. C. Johnson, K. M. Agesa, K. A. Shackelford, D. Tsoi, D. R. Kievlan *et al.*, “Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the global burden of disease study,” *The Lancet*, vol. 395, no. 10219, pp. 200–211, Jan. 2020. [Online]. Available: https://doi.org/10.1016/s0140-6736(19)32989-7

[3] M. Singer, C. S. Deutschman, C. W. Seymour, M. Shankar- Hari, D. Annane, M. Bauer *et al.*, “The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3),” *JAMA*, vol. 315, no. 8, pp. 801–810, 02 2016. [Online]. Available: https://doi.org/10.1001/jama.2016.0287

[4] M. A. Reyna, C. S. Josef, R. Jeter, S. P. Shashikumar, M. B. Westover, S. Nemati *et al.*, “Early prediction of sepsis from clinical data: The physionet/computing in cardiology challenge 2019,” *Critical Care Medicine*, vol. 48, pp. 210 – 217, 2019. [Online]. Available: https://physionet.org/content/challenge-2019/1.0.0/

[5] E. Alsentzer, J. Murphy, W. Boag, W.-H. Weng, D. Jindi, T. Naumann, and M. McDermott, “Publicly available clinical BERT embeddings,” in *Proceedings of the 2nd Clinical Natural Language Processing Workshop*. Minneapolis, Minnesota, USA: Association for Computational Linguistics, Jun. 2019, pp. 72–78. [Online]. Available: https://www.aclweb.org/anthology/W19-1909

[6] C. M. Torio and R. M. Andrews, “National inpatient hospital costs: The most expensive conditions by payer, 2011,” 2013. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK169005/

[7] N. Kissoon, “Sepsis guideline implementation: benefits, pitfalls and possible solutions,” *Critical Care*, vol. 18, no. 2, Mar. 2014. [Online]. Available: https://doi.org/10.1186/cc13774

[8] R. C. Bone, R. A. Balk, F. B. Cerra, R. P. Dellinger, A. M. Fein, W. A. Knaus *et al.*, “Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis,” *Chest*, vol. 101, no. 6, pp. 1644–1655, 1992. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S001236921638415X

[9] J. L. Vincent, R. Moreno, J. Takala, S. Willatts, A. D. Mendonc¸a, H. Bruining *et al.*, “The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure,” *Intensive Care Medicine*, vol. 22, no. 7, pp. 707–710, Jul. 1996. [Online]. Available: https://doi.org/10.1007/bf01709751

[10] Y. Wang, Y. Zhao, R. Callcut, and L. Petzold, “Integrating physiological time series and clinical notes with transformer for early prediction of sepsis,” 2022.

References

[11] Z. Liu, L. Wu, and M. Hauskrecht, “Modeling clinical time series using gaussian process sequences,” in *Proceedings of the 2013 SIAM International Conference on Data Mining*. SIAM, 2013, pp. 623–631.

[12] Z. Liu and M. Hauskrecht, “Clinical time series prediction: Toward a hierarchical dynamical system framework,” *Artificial intelligence in medicine*, vol. 65, no. 1, pp. 5–18, 2015.

[13] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, and J. Kang, “Biobert: a pre-trained biomedical language representation model for biomedical text mining,” *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240,

2020.
[14] A. E. Johnson, T. J. Pollard, L. Shen, L.-w. H. Lehman,

M. Feng, M. Ghassemi *et al.*, “Mimic-iii, a freely accessible critical care database,” *Nature*, 2016. [Online]. Available: https://www.nature.com/articles/sdata201635

[15] T. J. Pollard, A. E. Johnson, J. D. Raffa, L. A. Celi, R. G. Mark, and O. Badawi, “The eicu collaborative research database, a freely available multi-center database for critical care research,” *Scientific data*, vol. 5, no. 1, pp. 1–13, 2018.

[16] D. J. Demers and D. Wachs, “Physiology, mean arterial pressure,” 2019. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/30855814/

[17] J. Pennington, R. Socher, and C. Manning, “GloVe: Global vectors for word representation,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1532–1543. [Online]. Available: https://aclanthology.org/D14-1162

[18] P. Lewis, M. Ott, J. Du, and V. Stoyanov, “Pretrained language models for biomedical and clinical tasks: Understanding and extending the state-of-the-art,” in *Proceedings of the 3rd Clinical Natural Language Processing Workshop*. Online: Association for Computational Linguistics, Nov. 2020, pp. 146–157. [Online]. Available: https://aclanthology.org/2020.clinicalnlp-1.17

[19] Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, Rubenfeld G, Kahn JM, Shankar-Hari M, Singer M, Deutschman CS, Escobar GJ, Angus DC. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016 Feb 23;315(8):762-74. doi: 10.1001/jama.2016.0288. Erratum in: JAMA. 2016 May 24-31;315(20):2237. PMID: 26903335; PMCID: PMC5433435.

[20] Patrick Lewis, Myle Ott, Jingfei Du, and Veselin Stoyanov. 2020. Pretrained Language Models for Biomedical and Clinical Tasks: Understanding and Extending the State-of-the-Art. In *Proceedings of the 3rd Clinical Natural Language Processing Workshop*, pages 146–157, Online. Association for Computational Linguistics.

[21] Xiang Dai, Ilias Chalkidis, Sune Darkner, and Desmond Elliott. 2022. Revisiting Transformer-based Models for Long Document Classification. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 7212–7230, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Thank you!