MEDINFO 2025 — Healthcare Smart × Medicine Deep
M.S. Househ et al. (Eds.)

doi:10.3233/SHTI250956

Time-Series Bert for Sepsis Detection: Uncovering Patient Trajectories Through Vital Sign Embeddings

Roupen MINASSIAN, Meera RADHAKRISHNAN[1] and Kun YU
University of Technology Sydney, Australia

Abstract. This study adapts BERT for vital sign time-series analysis in sepsis detection. Using MIMIC-III data, our model's embeddings reveal patient clusters that partition septic from non-septic cases while capturing physiological complexity through diagnosis count distributions. The BERT-based classifier achieves robust performance in both Precision-Recall Area Under Curve (PR AUC), measuring precision maintenance across recall thresholds, and Receiver Operating Characteristic Area Under Curve (ROC AUC), quantifying septic/non-septic case discrimination. Unsupervised learning reveals patient subgroups with distinct physiological profiles, highlighting transformer architectures' ability to extract meaningful patterns from medical time-series for enhanced sepsis monitoring.

Keywords. Time-Series BERT, Vital Sign Analysis, Sepsis Detection, Patient Trajectories, Unsupervised Learning, Medical Time-Series

1. Introduction

Sepsis is a life-threatening condition that continues to be a leading cause of mortality and morbidity in intensive care units (ICUs) worldwide [1]. Early detection is crucial for improving outcomes, yet the heterogeneous nature of sepsis and its subtle, evolving symptoms pose significant challenges. Artificial intelligence (AI) has emerged as a powerful tool for early sepsis prediction, but many existing models [2], [3] use traditional machine learning on complex datasets, while more recent transformer-based approaches still rely on combined laboratory and vital sign data, limiting their deployment in resource-constrained settings [4][5].

This study introduces a framework for sepsis detection that leverages a subset of core physiological signals—heart rate, respiratory rate, temperature, and oxygen saturation—readily available in clinical and wearable health monitoring environments. The framework is built around an adapted Bidirectional Encoder Representations from Transformers (BERT) model [4]; while traditionally used for natural language processing, is tailored in this work to process physiological time-series data. This innovation allows the model to capture intricate temporal dependencies and interactions within the data whilst requiring only a minimal feature set. While existing sepsis detection systems [2], [3] often operate as 'black boxes', providing predictions without

[1] Corresponding Author: Meera Radhakrishnan, meeralakshmi.radhakrishnan@uts.edu.au.

revealing the underlying physiological patterns, our BERT-based framework addresses this limitation through interpretable embeddings. This interpretability is crucial in critical care, where understanding the reasoning behind model predictions is essential for clinical decision-making. By using minimal yet impactful inputs and leveraging explainable AI techniques, this approach underscores the potential of AI-driven health monitoring to support proactive care across diverse settings.

2. Methods

2.1. Dataset and Preprocessing

The approach is validated on the MIMIC-III dataset, a publicly available ICU database containing time-series data of vital signs recorded at varying intervals [5]. Preprocessing focused on standardizing four key physiological parameters (heart rate, oxygen saturation, body temperature, and respiratory rate) and removing invalid entries outside clinically normal ranges. To ensure consistent analysis, measurements are aggregated into hourly intervals and aligned into 60-hour windows starting from the first relevant diagnosis. The analysis includes sepsis cases identified through comprehensive International Classification of Diseases, Ninth Revision (ICD-9) diagnostic codes including direct sepsis diagnoses, systemic inflammatory response syndrome (SIRS) [6], and related conditions.

2.2. AI-Driven Time-Series Modeling

Originally developed for natural language processing tasks, BERT is repurposed in this study to analyze continuous physiological time-series data. We adapt BERT for continuous numerical sequences through architectural modifications: patient data is organized as a multivariate time-series tensor with dimensions [batch_size, max_len, num_features], where each patient's sequential measurements form a matrix of 60-time steps across all physiological features. Unlike text data requiring tokenization, these numerical features are directly processed through a dense linear projection layer that maps them into BERT's hidden dimension space, followed by layer normalization. This allows self-attention mechanisms to capture both temporal patterns and cross-feature interactions while handling missing values through normalization and padding strategies. Deep embeddings are extracted directly from the final transformer layer using the sequence output before the classification head, preserving the rich contextual representations of patient trajectories. The methodology, shown in Figure 1, involves training a BERT binary classifier from which deep embeddings are extracted. These embeddings undergo dimensionality reduction using Uniform Manifold Approximation and Projection (UMAP) [7] and DBSCAN clustering, with a k-nearest neighbors algorithm translating the identified patterns to new test cases. This pipeline effectively captures temporal dependencies in physiological signals, enabling robust sepsis progression analysis despite incomplete data.

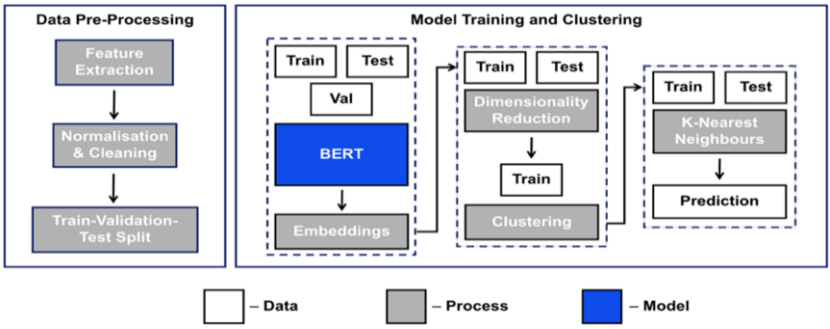

Figure 1. Pipeline for analyzing physiological time-series data using BERT embeddings, dimensionality reduction, and clustering. The model classifies cases via k-nearest neighbors based on established clusters.

3. Results

The model demonstrated strong discriminative performance with PR AUC of 0.89 and ROC AUC of 0.90. UMAP visualization of BERT-derived patient embeddings revealed distinct patient clusters based on health profiles (Figure 2). Most notably, sepsis patients formed several homogeneous subgroups in the embedding space, suggesting distinct pathophysiological phenotypes. These well-defined clusters of septic patients exhibited similar diagnostic profiles within groups while maintaining clear separation between groups, pointing to potential disease subtypes.

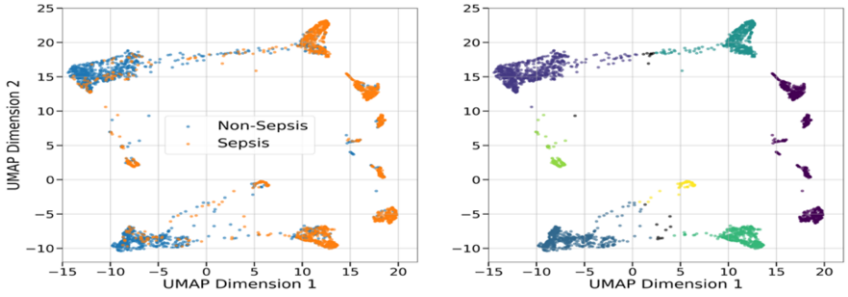

Figure 2. UMAP visualization of patient health records of test data. Left: Distribution of septic (orange) and non-septic (blue) patients. Right: DBSCAN-identified patient subgroups with noise points in black.

Further analysis of the diagnostic patterns between two major septic subgroups revealed significant differences in disease complexity (Figure 3). The first group (n=128) showed a higher burden of concurrent diagnoses (mean=2.51 ± 0.61, median=3.00), while the second larger group (n=205) exhibited a notably lower number of diagnoses per patient (mean=1.20 ± 0.54, median=1.00). This clear stratification in diagnostic burden suggests these computationally derived clusters may represent clinically meaningful sepsis phenotypes with varying comorbidities.

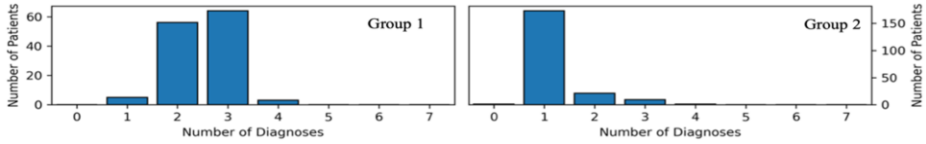

Figure 3. Distribution of the number of diagnoses per patient across two distinct septic patient groups. The histograms show the frequency of diagnoses per patient in each subgroup.

To further explore how these diagnostic patterns relate to patients' physiological states, we examined the relationship between vital sign measurements and diagnostic burden across the patient population. Figure 4 demonstrates that patients with similar vital sign patterns, as captured by BERT, tend to have similar numbers of diagnoses, suggesting a relationship between physiological state and diagnostic burden. The right panel focuses on a single cluster of a non-septic patient cohort, revealing a clear diagnostic gradient where the number of diagnoses systematically decreases from left to right, as confirmed by the median distribution plot below. This gradient structure indicates that even within a seemingly homogeneous patient group, there exists a continuous spectrum of diagnostic complexity that could inform clinical decision-making. The consistent diagnostic profiles within clusters and distinctions between septic and non-septic groups indicate that BERT embeddings capture meaningful disease patterns, warranting further study.

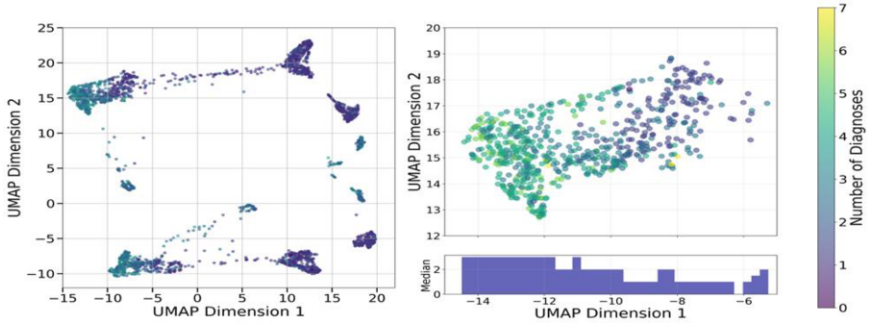

Figure 4. UMAP visualization of diagnosis patterns across patients. Left: Global distribution of diagnosis counts across all patients, with color intensity indicating the number of diagnoses per patient. Right: Focused analysis of a single cluster showing the diagnosis count gradient, with the lower panel displaying the median number of diagnoses across the horizontal axis, revealing local patterns in diagnosis density.

4. Discussion

Our BERT-based framework for vital sign analysis highlights deep learning's ability to uncover clinically meaningful patterns in physiological data through unsupervised analysis. The emergence of distinct patient phenotypes suggests that temporal vital sign patterns encode rich disease-related information often missed by conventional monitoring approaches. The strong alignment between physiological patterns and diagnostic profiles indicates that vital sign dynamics could serve as early indicators of disease complexity, enabling more proactive interventions. The relationship between cluster positioning and diagnostic burden suggests BERT embeddings capture broader

physiological states beyond sepsis-specific patterns. Systematic variations in diagnostic counts within clusters indicate that patients with similar vital sign patterns exhibit corresponding levels of complexity. The clear stratification between groups, with one cluster showing consistently higher comorbidities, suggests these computationally derived phenotypes reflect distinct disease manifestations. This multidimensional view of patient state could enable more nuanced clinical monitoring than traditional binary classification. However, several limitations remain. The generalizability of these clusters require validation across diverse populations and settings. While statistically distinct, the identified subgroups need clinical validation to confirm their practical significance. Additionally, though vital sign patterns correlate with diagnostic burden, causality cannot be inferred, as other factors like disease severity or care unit practices may contribute. Prospective studies should assess whether cluster-based monitoring improves outcomes compared to standard approaches. Future research should explore the clinical implications of these phenotypes and their impact on treatment strategies.

5. Conclusion

This study demonstrates how time-series BERT embeddings can uncover meaningful patterns in vital signs, extending beyond sepsis detection to reveal relationships between physiological patterns and patient complexity. The framework identifies interpretable patient subgroups while maintaining strong predictive performance, advancing nuanced patient monitoring. The discovered relationship between vital sign patterns and diagnostic burden suggests applications in risk stratification and resource allocation. Future work should validate these findings and integrate them into clinical decision support systems for more targeted sepsis management.

References

[1] Fleischmann-Struzek C, Rudd K. Challenges of assessing the burden of sepsis. Medizinische Klinik-Intensivmedizin und Notfallmedizin. 2023 Dec;118(Suppl 2):68-74, doi: 10.1007/s00063-023-01088-7.

[2] Yang M, Liu C, Wang X, Li Y, Gao H, Liu X, Li J. An explainable artificial intelligence predictor for early detection of sepsis. Critical care medicine. 2020 Nov 1;48(11):e1091-6, doi: 10.1097/CCM.0000000000004550.

[3] Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, Shimabukuro D, Chettipally U, Feldman MD, Barton C, Wales DJ. Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. JMIR medical informatics. 2016 Sep 30;4(3):e5909, doi: 10.2196/medinform.5909.

[4] Kenton JD, Toutanova LK. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of naacL-HLT 2019 Jun 2 (Vol. 1, p. 2), doi: 10.18653/v1/N19-1423.

[5] Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, Moody B, Szolovits P, Anthony Celi L, Mark RG. MIMIC-III, a freely accessible critical care database. Scientific data. 2016 May 24;3(1):1-9, doi: 10.1038/sdata.2016.35 (2016)

[6] Dellinger RP, Carlet JM, Masur H, Gerlach H, Calandra T, Cohen J, Gea-Banacloche J, Keh D, Marshall JC, Parker MM, Ramsay G. Surviving Sepsis Campaign guidelines for management of severe sepsis and septic shock. Critical care medicine. 2004 Mar 1;32(3):858-73, doi: 10.1097/01.CCM.0000117317.18092.E4.

[7] McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint :1802.03426. 2018 Feb 9. Available from: https://arxiv.org/pdf/1802.03426

[8] Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. Inkdd 1996 Aug 2 (Vol. 96, No. 34, pp. 226-231), doi:10.5555/3001460.3001507