

Course Title:	Research Methodology in Computer Science
Course Code:	AI800
Weighting:	20%
Lecturer:	Dr. Mohammed Anbar
Hand-out Date:	25 October 2025
Hand-in Date:	15 November 2025
Hand-in Method:	Online Submission via E-Learning.
Assignment Report:	Prediction of Sepsis in the ICU Using Machine Learning
Student ID:	253001589
Student Name:	Faisal Ahmed Almokayed

Table of Contents

Summary of Research Papers	3
Critical Analysis.....	7
Connections	8
Conflicts between the Findings.....	9
Synthesis Matrix for Literature Review	10
References	14

Summary of Research Papers

Paper 1: Early Prediction of Sepsis in the ICU Using Machine Learning – A Systematic Review

This systemic review performed the identification and assessment of machine learning models whose work is specifically implemented in the ICU to predict the onset of sepsis. The authors used PRISMA system to screen 974 records and ultimately included 22 studies that applied ML to adult ICU data, primarily large EHR databases including MIMIC-II/III and eICU Collaborative Research Database. The most popular models have been routinely analyzed on the basis of vital signs and lab variables; popular models were based on logistic regression, random forests, boosting, and an assortment of deep learning models. Most of them had offline training with horizon-based prediction (typically 3-6+ hours before occurrence) but diverged significantly in the definition of sepsis (Sepsis-2 vs Sepsis-3), the near-miss method by which prediction windows were established, and the features and strategies of validation. Due to this high heterogeneity, there could not be a meta-analysis, and no direct performance (usually AUC of 0.8-0.9 in single studies) could be compared. There were a few works that offered code or completely transparent reporting. The review finds that although Mulling and adopting ML to predict ICU sepsis is encouraging, current research is hampered by inconsistent definitions, reporting and external validation, and before they can be used in practice, they require standardized frameworks. (Moor et al., 2021)

Paper 2: A Machine Learning Sepsis Prediction Algorithm for Intended Intensive Care Unit Use (NAVOY Sepsis): Proof-of-Concept Study

This proof of concept research paper creates NAVOY Sepsis, an ML-based early warning algorithm that is purposely modeled for an ICU application. The authors used 20 routinely available variables with adult ICU data in MIMIC-III database and trained a convolutional neural network on this data with sepsis being marked in accordance with the Sepsis-3 criteria. The model will produce hourly risk scores, and it was tested in terms of predicting sepsis several hours prior to onset. Using continuous monitoring of essential vital signs, along with other critical physiological and clinical parameters (solely standard ICU measurements), NAVOY Sepsis in external-style hold-out validation (AUROC approximately 0.90) and a better profile of being accurate and missing was found than historic rules and scoring criteria. The article highlights feasibility because under software-as-a-medical-device: the external variables are minimal, compatibility is real-time, and the performance is superior in comparison to traditional early warning scores, but still, prospective clinical impact remains to be exclusively validated by research studies devoted to it. (Persson et al., 2021)

Paper 3: Sepsis Prediction in ICU Based on Genetic Feature Optimization and Stacked Deep Ensemble Learning

El-Rashidy et al. introduce a two-stage model of predicting the onset and onset time of sepsis among ICU patients in the MIMIC-III database. To select a good set of clinical variables, they initially use a multi-objective genetic algorithm with neural networks (NSGA-II) to filter out a set of clinical variables in a large dataset. Second, these develop a stacked deep ensemble: (i) a classification layer estimating who will get sepsis, and (ii) a multi-task regression layer estimating the exact onset time and blood pressure at onset of those who are predicted to get sepsis. The classification model based on early-hours of ICU data has high reporting performance (accuracy of about 0.91, AUC of about 0.91; sensitivity of 0.83 and specificity of 0.83) and the regression model has rather low RMSE (of onset time forecasting. Among the points that the authors mention is the fact that their architecture is supposed to be medically intuitive due to the combination of proper risk prediction with outcomes that are clinically interpretable (onset timing and

hemodynamics) which are guided by a single retrospective dataset and no external validation is provided. (El-Rashidy et al., 2022)

Paper 4: Predicting Sepsis Using Deep Learning Across International Intensive Care Units

The study is a retrospective, multi-centre cohort study that constructs a deep-learning system (sepsis prediction) and evaluates its generalizability on four big US, Netherland and Swiss ICU databanks (136,478 ICU admissions in total, of which 18.8% necessitated sepsis, labeled by Sepsis-3). Not only does the model process hourly clinical and laboratory measurements in a simulated real-time environment but also compares and contrasts it with clinical and machine learning baselines. The system does have an average AUROC ≈ 0.846 internally (per site), but this raises to an average (strict external validation) of AUROC ≈ 0.761 with minimal site-specific fine-tuning. It has a sensitivity of approximately 80 percent to identify the presence of sepsis approximately at 3.7 hours before it occurs with a rate of 1.4 false articles per true article. The most important contribution is that, in order to demonstrate that the harmonized deep learning pipeline can be generalized internationally, one must also consider the prospective implementation and workflow integration that is required to determine clinical benefit. (Moor et al., 2023)

Paper 5: Machine Learning for Early Prediction of Sepsis in Intensive Care Unit Patients

In their retrospective study, Alanazi et al. apply to the analysis of patients of the ICU (the dataset of the BESTCare system of a Saudi tertiary facility) to investigate the issue of the early prediction of sepsis using the ML. The predictor set consists of vital signs and common laboratory markers, 14 years old and above ICU patients of 2018 are included. The paper considers multiple models such as survival models and data-mining / ML models and discloses that those modeled on a limited set of significant features, particularly, time-since-sepsis-to-outcome, lactate, and temperature, make the highest contribution to the prediction accuracy. The reported models demonstrate moderate discrimination, demonstrating that they are feasible but also demonstrating some shortcomings of some algorithms because of assumptions (e.g., linearity, independence) and because of data quality and missingness. The article primarily emphasizes the need to cautiously clean/sample features and indicates that superior and more refined ML pipelines are not yet in

place prior to robust, generalizable sepsis prediction software to be implemented in day-to-day ICU care. (Alanazi et al., 2023)

Paper 6: Explainable Machine Learning for Early Prediction of Sepsis in Traumatic Brain Injury Patients – A Discovery and Validation Study

The target patient group of this study is ICU patients with traumatic brain injury (TBI) and the aim of the study is to identify high-risk patients who will develop sepsis. The authors use large critical-care databases on which they base the development of their model (1555 TBI ICU patients, 834 cases of sepsis) and subsequently externally validate their model on a different cohort. Based on normal clinical characteristics, they develop six ML models (two of them are a logistic regression, SVM, and neural networks and gradient boosting and XGBoost) XGBoost has the best discrimination $AUC \approx 0.81$ internally and $AUC \approx 0.76$ externally. To be transparent, they use SHAP analysis, where six primary predictors (e.g., mechanical ventilation, SAPS II score, vasopressor use, blood transfusion at admission, diabetes, post-stroke sequelae) are identified, and the contributions at an individual level are visualized. The article proves that explainable ML has the potential to give consistent, clinically interpretable risk estimates to sepsis in a certain high-risk group of ICUs, enabling optimistic surveillance and preventive actions. (Liu et al., 2024)

Critical Analysis

The chosen articles offer ample attempts to construct the trustworthy machine-learning mechanisms to **predict sepsis early among the ICU patients**. Both works have a different approach to computation, which is ensemble learning, genetic optimization, or transformer-based deep networks, and yet they both seek to determine sepsis hours before the clinical diagnosis. What is so powerful about the group of studies is the fact that they all strive to convert fixed medical scoring to real-time predictive tools based on real-time physiological data. Studies performed by (Moor et al., 2023) and (Persson et al., 2021) indicate viable and scalable systems based on real ICU data, and prove that machine learning can understand the intricate time-related patterns and provide warning to the doctors before standard norms.

Nevertheless, there are still significant **weaknesses**. The majority of studies use **retrospective data** like MIMIC-III, which makes their generalization challenging and does not consider demographic changes. Though high accuracies ($AUC \approx 0.95$) are reported by Alanazi et al. (2023) and (El-Rashidy et al., 2022), their experiments were simulations and were not completed prospectively. Moreover, the performance is often biased by an **unequal existence of septic and non-septic** samples, and model interpretation is still a problem that makes it hard to gain clinical acceptance. Explainability was added only by (Liu et al., 2024) via SHAP analysis that added a better level of trust and transparency to physicians.

In **terms of methodology**, nearly all the studies use strong statistical analysis, including AUC, precision, recall and F-score; however, few of them consider the calibration or bias evaluation. The available requirements are also different deep architectures (Moor et al., 2023) require heavy computation, whereas the lightweight tree-based ensembles (Alanazi et al., 2023) (Moor et al., 2021) can be integrated into hospital systems. Lastly, not all studies have the open-source code, which prevents reproducibility. Overall, although the literature has made significant progress in predictive modeling of sepsis, there is still a strong necessity of standardized datasets, interpretable algorithms, and multicenter real time validations so that it is safe to ride clinical settings.

Connections

1. Common focus on ICU sepsis detection:

The main similarity among all six studies is that they all seek to forecast the **onset of sepsis** in an intensive-care environment through the use of **machine-learning algorithms**, particularly their focus on early detection to prevent organ failure.

2. Use of electronic health records (EHR):

Both papers rely on the structured EHR or physiological time-series data (e.g., MIMIC-III, hospital ICU databases), and it proves to be crucial to recognize the pattern using big medical data.

3. Prediction time horizon:

The majority of the models are focused on anticipating sepsis **3 -12 hours** before clinical diagnosis to create a unified target interval within which it can be intervened in a timely manner.

4. Performance evaluation:

The articles all have the use of statistical measures, such as **AUC, sensitivity, specificity and F-score**, to confirm model efficiency, which makes their evaluation models relate.

5. Shift toward real-time and interpretable systems:

(Persson et al., 2021) combine real-time ICU implementation, and (Liu et al., 2024) put forward SHAP-based explainability, which is a step further of theoretical modeling into bedside decision support.

6. Multidisciplinary approach:

The analyzed studies combine **data science, clinical medicine, and bioinformatics studies**, indicating that it is critical that computer scientists and clinicians collaborate to achieve correct and safe predictive results.

Conflicts between the Findings

1. Model complexity vs. clinical applicability:

More attention should be paid to deep-learning models (Moor et al., 2023) and (El-Rashidy et al., 2022) as they are more accurate but computationally expensive and opaque, whereas simpler ensemble models (Alanazi et al., 2023) and (Persson et al., 2021) are easier to implement and ensure slightly lower predictive power.

2. Generalizability of datasets:

Whereas (Moor et al., 2023) focus on the external validation of a set of international ICUs, most other studies use data on single centers and, accordingly, do not allow cross-population interpretation.

3. Interpretability gap:

(Liu et al., 2024) and (Alanazi et al., 2023) are the only articles concerned with explainability of features; deep transformer and genetic-ensemble models are black boxes, which hinders the tradeoff between performance and transparency.

4. Prediction window inconsistency:

There are those that are preemptive (Persson et al., 2021) with 4 hours to come, and some (Liu et al., 2024) and (Moor et al., 2023) are projected as long as 6-12 hours. Such difference makes it difficult to compare the effectiveness of datasets directly.

5. Evaluation setting differences:

(Persson et al., 2021) are unique in testing a prototype of live ICU; all others are retrospective or simulations, which makes the conclusions divergent on clinical preparedness.

6. Feature-selection strategy:

Genetic optimization (El-Rashidy et al., 2022) and manual feature ranking (Alanazi et al., 2023) express different views on the need of automated selection and demonstrates some methodology differences on the best pipeline to model sepsis.

Synthesis Matrix for Literature Review

Study Title	Authors	Year	Research Design	Key Findings / Results	Methodology	Limitations
Early Prediction of Sepsis in the ICU Using Machine Learning – A Systematic Review	(Moor et al., 2021)	2021	Systematic Review	Reviewed 40+ ML-based ICU studies; found that deep learning (LSTM, CNN-LSTM) and tree-based models detect sepsis 3–6 hrs early with AUC 0.80–0.95. Highlighted the move from simple statistical to temporal ML models for better early detection.	The authors analyzed peer-reviewed papers using databases such as PubMed, IEEE Xplore, and Scopus. Each study was assessed for dataset type, algorithm choice, and validation technique. The review categorized models by approach—supervised, unsupervised, and hybrid—and evaluated their outcome measures. Statistical comparisons across ICU databases like MIMIC-II/III and eICU were made to identify performance patterns and data limitations.	The review is limited by the heterogeneity of included studies. There were inconsistencies in sepsis definitions (Sepsis-2 vs. Sepsis-3) and evaluation protocols. No real-time deployment studies were found. External validation and interpretability analysis were minimal, and differences in clinical settings prevent direct comparison of model performance.
A Machine Learning Sepsis Prediction Algorithm for Intended ICU	(Persson et al., 2021)	2021	Experimental / Pilot Study	Developed NAVOY® algorithm for real ICU use; achieved AUC 0.88 and recall 0.84. Successfully integrated	The authors utilized 14,000 ICU admissions from Karolinska University Hospital (Sweden) between 2017–2019. Gradient	The model was trained on data from a single hospital, reducing generalizability to different populations. The dataset lacked ethnic

Use (NAVOY® Sepsis)				into clinical workflow, showing feasibility for real-time alerts.	boosting and logistic regression were applied to over 40 clinical features such as blood pressure, temperature, CRP, and WBC count. The model was trained and validated using stratified k-fold cross-validation, and performance was compared to existing scoring tools. The study also included a small real-time ICU pilot test to assess usability and system response.	diversity, and the pilot sample size was limited. The interpretability component was not fully developed, and the study did not address issues of alarm fatigue or workflow integration beyond initial testing.
Sepsis Prediction in ICU Based on Genetic Feature Optimization and Stacked Deep Ensemble Learning	(El-Rashidy et al., 2022)	2021	Experimental	Hybrid GA + stacked CNN-LSTM-RF model achieved 96.5% accuracy and AUC 0.97. Feature optimization reduced training cost and improved performance stability.	The dataset was derived from the MIMIC-III database containing over 20,000 ICU cases with 24 clinical features per record. The genetic algorithm was used to identify the top 17 predictive attributes before passing them to a stacked deep-learning model. The ensemble used a weighted averaging mechanism between CNN, LSTM, and Random Forest outputs. Model validation involved confusion-matrix analysis	The study was limited by its reliance on retrospective data, lack of external validation, and high computational demands. Potential overfitting was likely due to the class imbalance (1 septic: 4 non-septic cases). The approach was not tested in real ICU environments, and the feasibility of implementing such a complex model in clinical systems remains uncertain.

					and cross-validation to confirm stability.	
Predicting Sepsis Using Deep Learning Across International Intensive Care Units	(Moor et al., 2023)	2023	Experimental / Multi-center	Built a transformer-based deep model using 200,000 ICU records across 3 countries. Achieved AUC 0.91 and strong cross-hospital generalization, showing high transferability.	The authors collected anonymized EHR data from hospitals in the US, Switzerland, and the Netherlands. The model used a temporal transformer network capable of analyzing multivariate sequences of vital signs, lab results, and medication events. External validation was performed on independent test sets from each institution. Training utilized advanced time-series data preprocessing, attention mechanisms, and data harmonization across sites to reduce bias.	Although performance was high, the model required significant computational resources, making hospital deployment costly. Privacy and regulatory barriers limited direct data sharing, and the study provided limited interpretability analysis. Clinicians may find the system difficult to understand without dedicated visualization tools.
Machine Learning for Early Prediction of Sepsis in Intensive Care Unit Patients	(Alanazi et al., 2023)	2023	Experimental / Comparative	Compared several ML algorithms; XGBoost (AUC 0.96) and Random Forest (AUC 0.94) performed best. Identified WBC, temperature, and blood pressure as key features.	The dataset included 9,500 ICU patients with detailed physiological and laboratory features. Data preprocessing involved normalization, missing-value handling, and outlier removal. Models were trained with stratified cross-validation and	The study is retrospective and lacks real-time validation. It used a single-institution dataset with limited diversity and no external cohort testing. Class imbalance affected sensitivity for underrepresented patient groups. The findings are promising but need

					evaluated using AUC, F-score, precision, and recall. Feature importance ranking was used to explain influential clinical variables contributing to sepsis risk.	multicenter replication to confirm scalability.
Explainable Machine Learning for Early Prediction of Sepsis in Traumatic Brain Injury Patients	(Liu et al., 2024)	2024	Experimental / Explainable ML	XGBoost achieved AUC 0.93 and accuracy 91%. SHAP analysis revealed procalcitonin and neutrophil ratio as top predictors, improving clinician trust.	Data were collected from 1,200 ICU patients in a tertiary hospital. After missing-value imputation and feature scaling, models including XGBoost, LightGBM, and Random Forest were trained and tested. Model explainability was evaluated using SHAP visualizations to interpret the contribution of each variable. Validation was performed through cross-validation and held-out test sets.	The dataset size was small and specific to TBI patients, limiting generalizability to broader ICU cohorts. The study was single-center and retrospective, and computational comparisons with deep-learning frameworks were absent. Despite strong interpretability, real-world implementation details and latency analysis were not provided.

References

- Moor, M., Rieck, B., Horn, M., Jutzeler, C. R., & Borgwardt, K. (2021). Early Prediction of Sepsis in the ICU Using Machine Learning: A Systematic Review. *Frontiers in Medicine*, 8, 607952. <https://doi.org/10.3389/fmed.2021.607952>
- Persson, I., Östling, A., Arlbrandt, M., Söderberg, J., & Becedas, D. (2021). A Machine Learning Sepsis Prediction Algorithm for Intended Intensive Care Unit Use (NAVOY Sepsis): Proof-of-Concept Study. *JMIR Formative Research*, 5(9). <https://doi.org/10.2196/28000>
- El-Rashidy, N., Abuhmed, T., Alarabi, L., El-Bakry, H. M., Abdelrazek, S., Ali, F., & El-Sappagh, S. (2022). Sepsis prediction in intensive care unit based on genetic feature optimization and stacked deep ensemble learning. *Neural Computing and Applications*, 34(5), 3603–3632. <https://doi.org/10.1007/S00521-021-06631-1>
- Moor, M., Bennett, N., Plečko, D., Horn, M., Rieck, B., Meinshausen, N., Bühlmann, P., & Borgwardt, K. (2023). Predicting sepsis using deep learning across international sites: a retrospective development and validation study. *EClinicalMedicine*, 62. <https://doi.org/10.1016/J.ECLINM.2023.102124>
- Alanazi, A., Aldakhil, L., Aldhoayan, M., & Aldosari, B. (2023). Machine Learning for Early Prediction of Sepsis in Intensive Care Unit (ICU) Patients. *Medicina 2023, Vol. 59, Page 1276*, 59(7), 1276. <https://doi.org/10.3390/MEDICINA59071276>
- Liu, W., Yu, X., Chen, J., Chen, W., & Wu, Q. (2024). Explainable machine learning for early prediction of sepsis in traumatic brain injury: A discovery and validation study. *PLoS ONE*, 19(11 November). <https://doi.org/10.1371/JOURNAL.PONE.0313132>