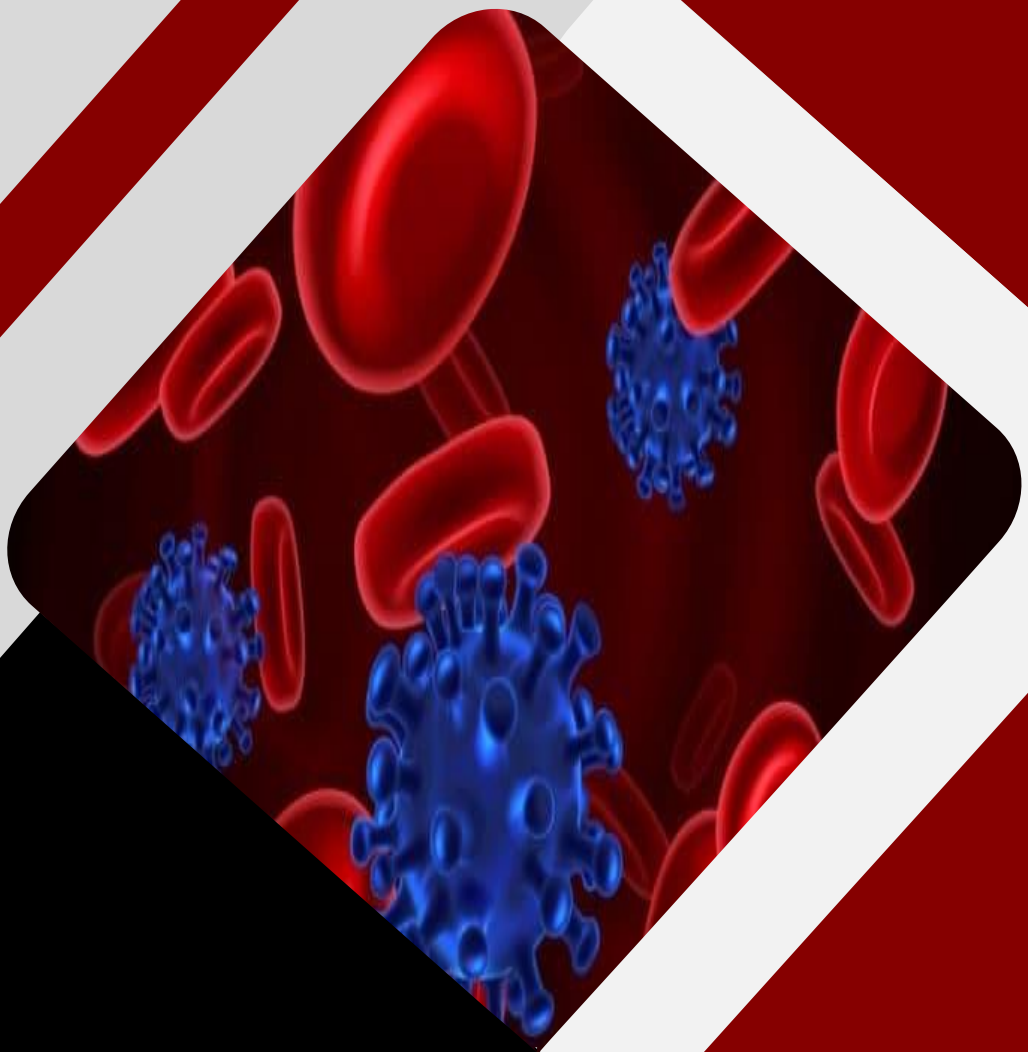

Statistical Signal Processing Assisted by Artificial Intelligence for Early Sepsis Detection Using Multimodal ICU Data

Prepared by: Dalal Qalajo

Introduction

Definition of Sepsis

First Definition: Introduced in 1991, based on the concept of SIRS, defined as a confirmed or suspected infection **plus the presence of two or more SIRS criteria**.

The four SIRS criteria:

1. Temperature above 38°C or below 36°C
2. Heart rate above 90 beats per minute
3. Respiratory rate above 20 breaths per minute
4. White blood cell count above 12,000 or below 4,000, or more than 10% immature cells

Problem: SIRS is very sensitive but not specific, as many patients meet the criteria, but it does not necessarily indicate sepsis.

Second Definition: An update to the first definition, published in 2001, which added:

- **Severe sepsis:** Sepsis plus organ dysfunction
- **Septic shock:** Sepsis plus hypotension unresponsive to intravenous fluids, requiring vasopressors

Third Definition: Published in 2016, it radically changed the first definition. Sepsis is now defined as **a life-threatening organ dysfunction caused by a dysregulated host response to infection.**

Dysregulated host response means:
The immune system releases large amounts of inflammatory chemicals into the bloodstream, not just locally and in limited quantities. This leads to widespread vasodilation, resulting in hypotension and reduced organ perfusion. Normally, vessels dilate slightly to allow blood and

immune cells to reach the infection site, and sometimes blood clots form locally to contain the infection. Abnormal systemic clotting, however, can block capillaries, stop blood flow to organs, and cause fluid leakage from vessels, reducing effective blood volume. Over time, organs receive less oxygen, leading to organ failure.

The SOFA score measures organ dysfunction associated with sepsis, evaluating six physiological systems:

1. **Respiration (lungs):** via blood oxygen levels
2. **Coagulation (blood):** via platelet count
3. **Liver:** via bilirubin levels
4. **Cardiovascular:** via blood pressure
5. **Kidneys:** via creatinine or urine output
6. **Nervous system:** via Glasgow Coma Scale

qSOFA is a rapid scoring system assessing only three parameters: respiration, blood pressure, and consciousness.

Definition of Septic Shock:

1. When blood vessels dilate excessively due to inflammation, causing hypotension unresponsive to intravenous fluids. The patient requires vasopressors to maintain a mean arterial pressure ≥65 mmHg, the minimum needed to ensure organ perfusion.
2. When insufficient blood reaches organs, cells switch to anaerobic metabolism, producing lactic acid. If lactate exceeds 2 mmol/L despite adequate fluid resuscitation, the patient is in septic shock.

Challenge of Early Sepsis Detection:

Early recognition and treatment of sepsis is a major global challenge because **each hour of delay increases mortality by 4–8%**. Current diagnostic tools are slow; for example, blood cultures to identify bacteria take 48 hours, while the patient's condition can deteriorate much faster.

Additionally, intensive care units generate massive amounts of multimodal data from vital signs, lab tests, and monitoring devices, overwhelming the capacity for manual analysis. This opens the door for advanced statistical signal processing and artificial intelligence.

However, key questions remain regarding the limits of early detection and the integration of statistical signal processing with AI. Addressing these questions offers an opportunity not only to develop predictive modeling but, more importantly, to **save lives through early sepsis detection and faster treatment**. This work represents my attempt to address these challenges.

Research Methodology

A **directed narrative review** methodology was adopted to select studies relevant to early sepsis detection using **biosignal processing and artificial intelligence** techniques. The selection process was based on the following criteria:

1. **Time frame:** Focus was placed on studies published between **2017 and 2021**, and beyond (up to 2025 when applicable), to ensure the use of up-to-date methods aligned with developments in machine learning and deep learning.
2. **Methodological value:** Studies that offered **innovative or pioneering approaches** in the field were included, such as:

- **Futoma et al. (2017):** First to present RNN with Gaussian Processes for early sepsis prediction.
- **Moor et al. (2021):** Large-scale international study validating DL models across multiple sites.
- **Kwon et al. (2021):** Deep model using **ECG only**.
- **Gupta et al. (2020):** Use of multiple physiological signals with machine learning.
- **Yamanashi et al. (2021):** Introduction of EEG (BSEEG score) as an early indicator of sepsis-related mortality.

3. **Accessibility and reliability:** Only **fully available studies** published in **peer-reviewed journals or conferences** (e.g., ICML, Lancet Digital Health, Scientific Reports, etc.) were considered.

Summary of Studies

The studies will be presented in a way that highlights the **challenges overcome at each stage of the work**, providing insight into what was done, how it was accomplished, and the key takeaways from reviewing these five studies.

Study 1: Learning to Detect Sepsis Using Multi-Task Gaussian Process with RNN

Step 1: Data Collection

Challenge: Detecting sepsis is difficult because its symptoms resemble those of other diseases.

Work: Collected a large dataset of **51,697 cases over 18 months**, with diversity across:

8 medications, 29 diseases, 3 types of hospital admission (transfer, urgent, emergency), 6 vital signs, 28 laboratory values, 3 demographic variables

This diversity allows the model to learn **subtle patterns that distinguish sepsis from other conditions**.

Step 2: Data Preprocessing

Challenge: Data is irregular, incomplete, and noisy.

Work: Used a **Multi-Output Gaussian Process algorithm**, which assumes the data comes from an unknown continuous function and estimates it probabilistically (expected value + confidence interval).

Key steps:

Filling missing values: For example, if blood pressure is measured at 3 AM and 5 AM but missing at 4 AM, the algorithm estimates a value with a confidence interval.

Noise removal: By modeling the underlying function, random disturbances are filtered out.

Data regularization: Converts data into a **structured time grid (hourly)**. Example: lactate measured every 8 hours is interpolated to provide hourly values.

Step 3: Modeling

Challenge: Traditional methods (SOFA, qSOFA, SIRS, MWES, NWES) ignore complex temporal relationships and use only a few variables.

Work: Trained an **LSTM model** capable of handling complex temporal relationships, such as the correlation between today's temperature and blood pressure measured four days earlier.

Challenge: Accurately identifying the onset time of sepsis.

Work: Added a **temporal window** for training: instead of training on a single onset moment, the model was trained on a window from 2 hours before to 6 hours after onset. This reduces the impact of minor errors in pinpointing the exact start time.

Step 4: Training

Challenge: Overfitting the model.

Work: Applied **L2 regularization** to prevent the model from relying on large individual weights by penalizing the sum of squared weights, encouraging balanced smaller weights.

Early stopping was used to **halt training before overfitting occurs**, as L2 alone cannot fully prevent overfitting with prolonged training.

Results:

The model was evaluated using **Matched Lookback Validation**, testing hour by hour from the event onset to determine how far in advance sepsis could be predicted.

The model outperformed **traditional RNNs, MWES, NWES, logistic regression, and random forest models**.

It predicted sepsis **4 hours in advance** with **1.4 false alarms per true alarm at 80% sensitivity**, compared to 3.2 false alarms per true alarm using MWES.

Study 2: Predicting Sepsis Using Deep Learning Across International Sites: A Retrospective Development and Validation Study

Step 1: Data Collection

Challenge: Few multicenter studies exist, and transferring models between countries and hospitals is difficult due to differences in clinical practices, equipment, and patient populations.

Work: Used **four databases from three different countries** (USA, Netherlands, Switzerland) covering the period 2001–2016, including **136,478 ICU admissions**.

Challenge: Detecting sepsis is difficult because its symptoms resemble other conditions.

Work: Selected **59 variables** from vital signs and laboratory tests, plus **4 static variables** (age, gender, …).

Note: Medications were excluded because including treatment could be misleading (treatment indicates physician suspicion of sepsis), allowing the model to capture the **full and complex clinical picture**.

Step 2: Data Preprocessing

Challenge: Data was irregular, incomplete, and noisy.

Work: Standardized the data and handled missing values. No specific algorithms were mentioned, so standard programming and development environments were likely used.

Step 3: Modeling

Challenge: Traditional methods (SOFA, qSOFA, SIRS, MWES, NWES) ignore complex temporal relationships and use few variables.

Work: Trained a **Self-Attention Deep Network** capable of modeling **short- and long-term temporal relationships**, e.g., the relationship between today's temperature and blood pressure measured four days earlier.

Step 4: Training

Challenge: Difficulty transferring models between countries and hospitals.

Work:

Internal validation: Train and test on the same hospital database.

External validation: Train on one database, test on another (cross-country).

Results:

Internal validation: Accuracy 0.846; at 95% sensitivity, 1.4 false alarms per true alarm; predicted sepsis **3.7 hours in advance**.

External validation: Accuracy 0.76; at 95% sensitivity, 2.1 false alarms per true alarm; predicted sepsis **1.7 hours in advance**.

Improvements:

Ensemble strategy: Train the model separately on three databases and test on the fourth, repeating this process. For sepsis prediction, the **highest output among the three models** is taken (like "listen to the first warning from any model"). Performance improved or matched the best single-database model (unknown in advance), enhancing **generalization across hospitals and countries**.

Fine-tuning: Used 10% of the new hospital's test data to update the model before testing for adaptation. Resulted in **accuracy 0.807 at 95% sensitivity**, very close to internal validation performance.

Study 3: Deep Learning Model for Sepsis Screening Using Electrocardiography (ECG)

Step 1: Data Collection

Challenge: Traditional sepsis screening relies on vital signs (measured intermittently) and laboratory tests (requiring infrastructure, staff, and time). This makes real-time or remote detection impractical, especially in resource-limited settings and during epidemics.

Action: Researchers collected digitally stored ECG data with two demographic variables (age and gender) from **46,017 patients across two hospitals**. ECGs are non-invasive, inexpensive, continuous, and can be recorded with wearable devices. Using ECG as the primary input laid the foundation for a scalable, real-time screening tool.

Step 2: Data Preprocessing

Challenge: Raw ECG signals are noisy and contain distortions (baseline drift, vector noise, motion artifacts), which can obscure subtle sepsis-related patterns.

Action: Applied a band-pass filter to reduce noise. Removed unstable segments (first/last seconds).

Normalized signals (standardization) since ECGs were not on the same scale (some stronger/weaker, shifted up/down due to device differences or body habitus). Normalization

ensured the neural network focused on **temporal and morphological features (QRS, QT, axes, etc.)**, rather than just signal amplitude.

Performed **data augmentation** by adding synthetic noise (e.g., baseline shifts, small oscillations simulating patient movement or device interference). The goal was **not to corrupt data**, but to teach the model robustness under real-world noisy conditions.

Step 3: Modeling

Challenge: Traditional statistical methods (e.g., logistic regression) cannot capture the subtle, nonlinear ECG abnormalities associated with sepsis.

Action: Built a **Residual Convolutional Neural Network (ResNet-like CNN)** capable of learning complex, high-dimensional patterns from ECG waveforms.

Skip connections (residual links) solved the **vanishing gradient problem**: if intermediate layers fail to learn useful features, the raw signal can still flow forward, and gradients can flow backward efficiently, improving learning in deep networks.

Step 4: Training

Challenge: Models often fail when transferred between hospitals.

Action: The model was trained and validated on data from one hospital, then tested on another, demonstrating **generalizability** and reducing the risk of bias.

Results:

Sepsis detection accuracy: 0.901 (internal validation), 0.863 (external validation).

Septic shock detection: 0.906 (internal), 0.899 (external).

Single-lead ECG (wearable-compatible): accuracy 0.845.

Thus, researchers showed that a deep learning model can detect sepsis **even from simple, single-lead ECG signals**, paving the way for at-home and wearable-based sepsis screening rather than relying solely on hospital-grade equipment.

Study 4: Early Detection of Sepsis-Induced Deterioration Using Machine Learning

Step 1: Data Collection

Challenge: Traditional detection of sepsis relies on discrete vital signs and non-specific symptoms (such as fever, tachycardia), which are not definitive. By using continuous high-frequency physiological signals, the study captured subtle physiological changes that static measurements miss.

Action: Data were collected from 132 patients suspected of sepsis in the emergency department during the first 48 hours of hospital admission. Continuous signals were recorded: ECG (500 Hz), plethysmography/SpO_2 (125 Hz), and respiratory rate (62.5 Hz). Outcomes (ICU transfer, organ failure, death) were used to classify patients as deteriorating or not.

Step 2: Data Preprocessing

Challenge: Sepsis-related signals are noisy and incomplete, making preprocessing difficult.

Action: The study mitigated this by segmenting signals into short, artifact-free 5-minute windows and applying precise heartbeat detection from R-peaks, ensuring usable inputs despite missing data. Three strategies were tested to transform raw biosignals into meaningful features:

1. Histogram of derivatives: extracting first derivatives (rate of change) and second derivatives (acceleration of change) for each signal, and plotting histograms of how often certain changes occurred. The idea is to track not only heart rate but also how

fast and how steadily it changes. Differences between successive histograms were also calculated to capture evolving instability over time.

2. Temporal comparison: comparing consecutive signal segments to detect whether patterns became more disordered as time progressed.
3. Wavelet transform with autoregressive modeling: decomposing ECG signals into frequency and temporal components (e.g., QRS, T waves) and extracting features. Instead of just looking at average heart rate, this approach examined whether the morphology of individual beats changed (e.g., alterations in QRS or T waves), which may appear early before overt deterioration.

Step 3: Modeling and Training

Challenge: Ensuring model generalization across patients rather than just fitting one individual's data.

Action: Machine learning models (SVM, Random Forest, GBM, kNN, MLP, Naïve Bayes, Logistic Regression) were trained on 90% of the data and tested on 10%, with cross-validation applied.

Results:

The first two strategies performed poorly, with 56% accuracy. The wavelet plus autoregressive strategy achieved the best result: 62.4% accuracy with an SVM. The main challenges were limited signal diversity and small dataset size. Nonetheless, the study showed that ECG morphological features provide effective predictive power and could approach the accuracy of traditional methods, offering a new path for early sepsis detection.

Study 5: Mortality among patients with sepsis associated with a bispectral electroencephalography

BSEEG (Bispectral EEG) is an advanced method of analyzing EEG signals. Instead of looking only at raw waveforms, it applies bispectral analysis, a mathematical method that captures interactions between different frequency components of brain signals.

Step 1: Data Collection

Action: Researchers collected raw EEG data from 628 patients at the same hospital, each recording lasting 3–10 minutes, along with demographic data (age, gender, comorbidities).

Step 2: Feature Engineering

Challenge: Current evaluation of sepsis-associated encephalopathy (SAE) relies on subjective neurological exams (e.g., Glasgow Coma Scale), which vary among physicians. An objective, quantitative biomarker for brain injury is needed.

Action: Raw EEG signals were processed using a pre-existing algorithm. The BSEEG score is based on the ratio of low-frequency (slow) to high-frequency (fast) brainwave power. An increase in this ratio indicates abnormal brain slowing, a marker of encephalopathy. The ratio was normalized into a standardized score.

Step 3: Patient Stratification – From Simple Diagnosis to Risk Assessment

Challenge: While sepsis diagnosis is important, not all septic patients face the same mortality risk. Prioritization requires identifying those most at risk of deterioration.

Action: Patients were divided into groups based on two factors: (1) clinical sepsis status from medical records, and (2) BSEEG score (high or low). This produced four distinct groups for comparison (e.g., “Sepsis + High BSEEG” vs. “Sepsis + Low BSEEG”).

Outcome: This classification allowed researchers to test whether BSEEG could stratify mortality risk within the sepsis patient group.

Step 4: Data Analysis – Statistically Validating BSEEG

Challenge: A new tool must provide additional information beyond age and comorbidities.

Action: Researchers used Cox regression to analyze the relationship between BSEEG score and mortality, adjusting for age, gender, and comorbidities.

Result: BSEEG was found to be an independent predictor of mortality. Even with two patients of the same age and comorbidity level, the one with a higher BSEEG score had a greater risk of death.

Results:

As a diagnostic tool for confirming sepsis, BSEEG was not accurate.

As a prognostic tool, it was excellent: septic patients with high BSEEG scores had much higher mortality. For example, the “Sepsis + High BSEEG” group had the worst survival rate (only 46.3% alive at one year). The risk increased in a dose-dependent manner: higher scores correlated with worse outcomes.

Objective risk stratification was confirmed, showing that BSEEG predicts mortality independently of traditional risk factors.