

---

## Master Thesis

---

Max Mustermann

# **Combining Machine-Learning and Dynamic Network Models to Improve Sepsis Prediction**

January 30, 2026

---

supervised by:

Albert Dreistein

Wolfgang St. Pauli

---

Hamburg University of Technology  
Institute for Biomedical Imaging  
Lottestraße 55  
22529 Hamburg

University Medical Center Hamburg-Eppendorf  
Section for Biomedical Imaging  
Martinistraße  
20246 Hamburg



Hiermit erkläre ich, Author, an Eides statt, dass ich die vorliegende Masterarbeit im Studiengang Example-Program selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel – insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen – benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht. Ich versichere weiterhin, dass ich die Arbeit vorher nicht in einem anderen Prüfungsverfahren eingereicht habe. Sofern im Zuge der Erstellung der vorliegenden Abschlussarbeit gKI-basierte elektronische Hilfsmittel verwendet wurden, versichere ich, dass meine eigene Leistung im Vordergrund stand und dass eine vollständige Dokumentation aller verwendeten Hilfsmittel gemäß der Guten Wissenschaftlichen Praxis vorliegt. Ich trage die Verantwortung für eventuell durch die gKI generierte fehlerhafte oder verzerrte Inhalte, fehlerhafte Referenzen, Verstöße gegen das Datenschutz- und Urheberrecht oder Plagiate.

Hamburg, den 30.01.2026

(Author)

## **Abstract**

# Contents

<b>1. Introduction</b>	<b>3</b>
<b>2. Medical Background (Sepsis)</b>	<b>5</b>
2.1. Biology of Sepsis . . . . .	5
2.2. Sepsis-3 Definition . . . . .	8
2.3. Sepsis Prediction . . . . .	10
<b>3. State of the Art</b>	<b>13</b>
3.1. Model-Based Approaches . . . . .	13
3.2. Data-Driven Approaches . . . . .	13
3.3. Summary of State of the Art . . . . .	16
<b>4. Model Background (Dynamic Network Model)</b>	<b>17</b>
4.1. Theoretical Background: The Kuramoto Oscillator Model . . . . .	18
4.2. Description . . . . .	21
4.3. Implementation . . . . .	24
4.4. Limitations and Research Direction . . . . .	31
4.5. Summary of the Dynamic Network Model . . . . .	33
<b>5. Method (Latent Dynamics Model)</b>	<b>35</b>
5.1. Formalizing the Prediction Strategy . . . . .	35
5.2. Architecture . . . . .	39
5.3. Training Objective and Auxiliary Losses . . . . .	47
5.4. Inference . . . . .	49
5.5. Assessing the Prediction Performance . . . . .	49
5.6. Summary of Methods . . . . .	51
<b>6. Experiment</b>	<b>53</b>
6.1. Data . . . . .	53
6.2. Implementation Details . . . . .	56
6.3. Results . . . . .	59
<b>7. Discussion</b>	<b>67</b>
7.1. Addressing the Research Questions . . . . .	67
7.2. Key Strengths . . . . .	68
7.3. Limitations and Challenges . . . . .	68
7.4. Future Directions . . . . .	69
<b>A Appendix</b>	<b>73</b>
A.1 SOFA - Details . . . . .	73
A.2 DNM as Lie Formulation . . . . .	74
A.3 Input Concepts . . . . .	74

A.4 Latent Dynamic Model Architecture Parameters . . . . .	76
<b>Bibliography</b>	<b>77</b>

## List of tables

Table 4.1.	Dynamic Network Model Notation Table .....	22
Table 4.2.	Simulation Parameterization .....	25
Table 4.3.	Specific $\beta$ - $\sigma$ combinations to illustrate simulation results .....	28
Table 5.4.	Latent Dynamics Model Notation Table .....	39
Table 5.5.	Training Objectives .....	48
Table 6.6.	Characteristics and demographics of the cohort .....	55
Table 6.7.	Hyperparameters and learnable parameters .....	59
Table 6.8.	Performance Comparison to Baseline Model .....	62
Table 1.9.	TODO .....	73
Table 1.10.	Static input features for the prediction task .....	74
Table 1.11.	Dynamic input features for the prediction task. ....	74
Table 1.12.	Detailed parameter count of every module .....	76
Table 1.13.	.....	76

## List of figures

Figure 2.1.	Illustration and microscopic image of Stroma and Parenchymal Cells .....	6
Figure 2.2.	Timings of the Sepsis-3 definition .....	10
Figure 3.3.	Offline vs. Online prediction .....	14
Figure 4.4.	Synchronization transition in the Kuramoto model .....	19
Figure 4.5.	Schematic illustration of the Dynamic Network Model .....	21
Figure 4.6.	Dynamic Network Model Initialization .....	26
Figure 4.7.	Snapshots of simulated system states .....	29
Figure 4.8.	Temporal evolution of the phase- and frequency-synchronization metrics ..	30
Figure 4.9.	Phase Space of $\sigma$ and $\beta$ .....	31
Figure 5.10.	Infection Indicator Module Architecture .....	40
Figure 5.11.	SOFA Predictor Module Architecture .....	43
Figure 5.12.	Latent Lookup .....	45
Figure 5.13.	Decoder Architecture .....	46
Figure 5.14.	Complete Latent Dynamics Model Architecture .....	47
Figure 6.15.	Cohort selection and exclusion process. ....	54
Figure 6.16.	Implementation of the latent encoder and decoder module. ....	57
Figure 6.17.	Progression of training and validation losses. ....	60
Figure 6.18.	Receiver Operating Characteristic and Precision-Recall Curve. ....	61
Figure 6.19.	Receiver Operating Characteristic and Precision Recall Curve. ....	63
Figure 6.20.	Distribution of latent predictions. ....	64
Figure 6.21.	Individual patient trajectories demonstrate representations of clinical trends. ....	65

# Acronyms Index

<b>ABX:</b>	Antibiotics
<b>AUPRC:</b>	Area Under Precision Recall Curve
<b>AUROC:</b>	Area Under Receiver Operationg Curve
<b>BCE:</b>	Binary Cross Entropy
<b>DAMP:</b>	Damage-Associated Molecular Patterns
<b>DL:</b>	Deep Learning
<b>DNM:</b>	Dynamic Network Model
<b>EHR:</b>	Electronic Health Record
<b>FN:</b>	False Negatives
<b>FP:</b>	False Positives
<b>FPR:</b>	False Positive Rate
<b>FSQ:</b>	Finite Scalar Quantization
<b>GELU:</b>	Gaussian Error Linear Unit
<b>GPU:</b>	Graphics Processing Unit
<b>GRU:</b>	Gated Recurrent Unit
<b>ICU:</b>	Intensive Care Unit
<b>IQR:</b>	Inter Quantile Range
<b>JIT:</b>	Just In Time Compilation
<b>LDM:</b>	Latent Dynamics Model
<b>LOS:</b>	Length Of Stay
<b>LSTM:</b>	Long Short-Term Memory
<b>MIMIC:</b>	Medical Information Mart for Intensive Care
<b>ML:</b>	Machine Learning
<b>MLP:</b>	Multi Layer Perceptron
<b>MSE:</b>	Mean Squared Error
<b>ODE:</b>	Ordinary Differential Equation
<b>PAMP:</b>	Pathogen-Associated Molecular Patterns
<b>PID:</b>	Proportional-Integral-Derivative
<b>PINN:</b>	Physics Informed Neural Networks
<b>PRR:</b>	Pattern Recognition Receptors
<b>qSOFA:</b>	Quick Sequential Organ Failure Assessment
<b>RAG:</b>	Retrieval Augmented Generation
<b>ReLU:</b>	Rectified Linear Unit
<b>RNN:</b>	Recurrent Neural Networks
<b>SIRS:</b>	Systemic Inflammatory Response Syndrome
<b>SI:</b>	Suspected Infection
<b>SOFA:</b>	Sequential Organ Failure Assessment
<b>SVM:</b>	Support Vector Machine
<b>TN:</b>	True Negatives
<b>TPR:</b>	True Positive Rate
<b>TP:</b>	True Positives
<b>TUHH:</b>	Hamburg University of Technology
<b>YAIB:</b>	Yet Another ICU Benchmark

- fix Bibliography in TOC
- Sources misc + link
- TUHH address wrong in template

actual functional model what is learned connecting parts



# 1

## Introduction



# 2

## Medical Background (Sepsis)

As the most extreme course of an infectious disease, sepsis poses a very serious health threat, with a high mortality rate and frequent long-term consequences for survivors. In 2017, an estimated 48.9 million people worldwide suffered from sepsis and the same year, 11 million deaths were associated with sepsis [1], which makes up 19.7% of yearly deaths, making it the most common cause of in-hospital deaths. Even with successful treatment, around 40% of those affected suffer long-term consequences, such as cognitive, physical or physiological problems, the so called *post-sepsis syndrome* [2]. The observed risk of mortality significantly differs between lower to middle income countries with > 50% and high income countries with < 25%. Overall, septic diseases represent an enormous burden on the global healthcare system.

Quantitatively, almost half of all sepsis-related deaths occur as a secondary complication of an underlying injury or a non-communicable, also known as chronic disease [3]. But the underlying biological triggers for sepsis as well as individual patient progressions are highly diverse and heterogeneous. Moreover, a septic condition can not be reduced to a single specific physiological phenomenon, instead it combines multiple complex and interdependent processes across different biological scales.

The following chapter is structured around the most commonly used and widely accepted sepsis definition, which is referred to as *Sepsis-3*, it characterizes sepsis as a “life-threatening organ dysfunction caused by a dysregulated host response to infection” [4]. Starting with Section 2.1, both the pathology and underlying biology of that definition are described phenomenologically. Followed by Section 2.2, which provides a more medical and technical overview to this definition. In Section 2.3 the necessity for reliable and clinically practical sepsis prediction systems is discussed and how these systems are utilized in clinical practice.

### 2.1. Biology of Sepsis

This part tries to give an introduction into the biological phenomena that underlie sepsis. Starting with an explanation on how human tissue is reacting to local infections or injuries on a cellular level in Section 2.1.1 and how this can escalate to *cytokine storms* in Section 2.1.2 and ending with systemic organ failure in Section 2.1.3.

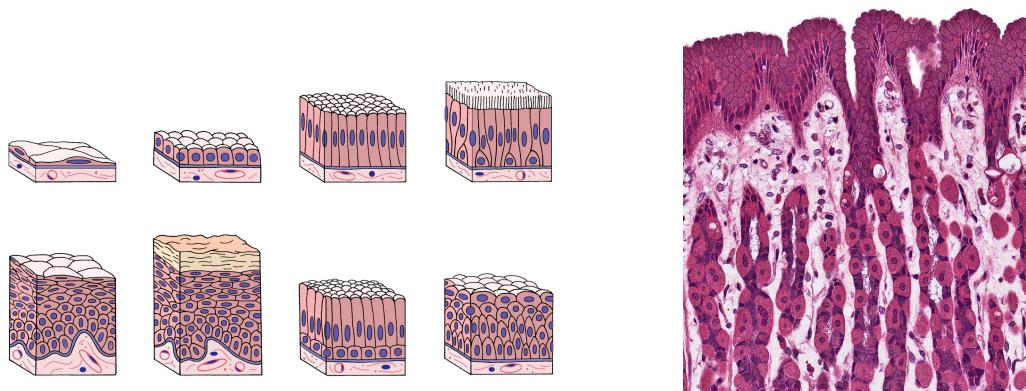
Certain details and specificities are left out when not essential for the understanding of this project. More detailed explanations can be found in the primary resources provided throughout this section.

### 2.1.1. Cellular Origins

Human organ tissue can be differentiated into two broad cell-families called *parenchymal* and *stroma* which are separated by a thin, specialized boundary layer known as the *basal lamina*.

The parenchymal cells perform the primary physiological functions of an organ, with every organ hosting distinct parenchymal cells. For example, Cardiomyocytes in the heart drive the contraction, relaxation and therefore blood pumping, Hepatocytes in the liver doing metabolism and detoxification and Neurons in the brain providing signal transmission [5].

Everything not providing organ-specific functionalities forms the stroma, that includes the structural or connective tissue, blood vessels and nerves. The stroma not only contributes to the tissues structure, but it also actively participates in biochemical signaling and immune regulation. This way it helps to maintain a healthy and balanced tissue, the *homeostasis*, and enables coordinated responses to injury or infection [6].



**Figure 2.1:** On the left hand side, an illustration of parenchymal cells sitting on top of the basal lamina (thin light-blue section) and the stroma in very light pinkish color building the base. The illustration shows the parenchymal cells in different shapes and organization [7]. On the right hand side a microscopic image of stomach tissue, with tightly packed parenchymal cells on the top and stroma cells as connective tissue in the light-pinkish color. The basal lamina is not visible at this level of magnification [8].

The pathogen summarizes all types of organisms that can be harmful to the body, this includes germs, fungi, algae, or parasites. When a pathogen enters the body through the skin, a mucous membrane or an open wound, the first line of nonspecific defense, the innate immune system [9], gets activated.

This rapid response does not require the body to have seen the specific pathogen before, opposed to a slower more specific and adapted immune response. Instead, the innate immune system can be triggered by sensing commonly shared features of pathogens, in case of germs known as Pathogen-Associated Molecular Patterns (PAMP), for injury called Damage-Associated Molecular Patterns (DAMP) [10]. The PAMP's and DAMP's can be detected by Pattern Recognition Receptors (PRR), which are found in resident immune cells, as well as stroma cells. Once a pathogen is detected a chain reaction inside the cell leads to the creation and release of signaling proteins called *cytokines* [11].

Cytokines are a diverse group of small signaling proteins which play a special role in the communication between other cells, both neighboring and across larger distances through the bloodstream. They are acting as molecular messengers that coordinate the recruitment of circulating immune cells and will guide them to the location of infection or injury [11].

Besides their role in immune activation where cytokines regulate the production of anti- and pro-inflammatory immune cells which help with the elimination of pathogens and trigger the healing process right after. They are also participating in the growing process of blood cells.

One specialty of these relatively simple proteins is that they can be produced by almost every other cell, with different cells being able to produce the same cytokine. Further, cytokines are redundant, meaning targeted cells can show identical responses to different cytokines [12], these features seem to fulfill some kind of safety mechanism to guarantee vital communication flow. After release cytokines have relatively a short half-life (only a few minutes) but through cascading-effects the cytokines can have substantial impact on their micro-environment.

### 2.1.2. Cytokine Storms

The host's dysregulated response to an infection connected to the septic condition is primarily driven by the excessive and uncontrolled release of cytokines and other mediators. Under normal circumstances, the release of inflammatory cytokines is tightly regulated in time and magnitude. After the pathogen is detected, the release of cytokines is quickly initiated. The release peaks as immune cells are recruited and automatically fades out once the initial pathogen is controlled and the host returns to a healthy and balanced state, the homeostasis.

In certain scenarios a disturbance to the regulatory mechanisms triggers positive inflammatory feedback loop, followed by a massive release of pro-inflammatory cytokines. These cells further activate additional immune and non-immune cells, which in turn amplify the cytokine production, creating a self-reinforcing cycle of immune activation [13]. This ultimately leads to a continuous and uncontrolled release of cytokines that fails to shut down. With this overreaction, called *cytokine storm*, the immune response and release of inflammatory mediators can damage the body more than the infection itself.

Although the quantity of cytokines roughly correlates with disease severity, concentrations of cytokines vary between patients, time and even different body-parts, making a distinction between an appropriate reaction and a harmful overreaction almost impossible [13]. Out of all cytokines, only a small subset or secondary markers can be measured through blood samples to detect increased cytokine activity. This limited accessibility makes cytokines difficult to study in general, they prove to be little useful as direct indicators of pathogenesis or diagnostic purposes.

Since the 90s there has been a lot of research focused on cytokines and their role in the innate immune system and overall activation behavior. Multiple therapeutic interventions have been tested in clinical trials, yet none have achieved a significant improvement in survival outcomes [10]. This emphasizes the complexity of sepsis as a systemic syndrome rather than a single-cause disease, and suggests that cytokine storms are an emergent property rather than the

result of any one molecular trigger. To this day, the fundamental principles that govern the transition from a regulated immune response to a self-destructive cytokine storm remain not fully understood.

### 2.1.3. Systemic Consequences and Organ Failure

While more and more cytokines are released, they flood not only infected areas, but also surrounding parts of the tissue and circulation, causing localized inflammatory response to become systemic. The widespread cytokine reaction starts to disrupt the normal metabolism of parenchymal cells in organs due to a deficiency in oxygen and nutrients.

To compensate, cells switch from their usual oxygen-based metabolism to an *anaerobic glycolysis* [14], generating energy less efficiently from glucose. As a result, metabolic by-products such as lactate accumulate making the surrounding environment more acidic, which further harms the cells and leads to more cellular dysfunction.

At the same time, the mitochondria start to fail. The walls of blood vessels become leaky, allowing fluids to move into surrounding tissue. This causes swelling and lowers the blood pressure, which in turn reduces the oxygen supply even further [10].

Step by step, the death of cells spreads throughout the body and affects organ functionality. When multiple organs fail simultaneously, the condition becomes irreversible [4]. A patient at this stage is in septic shock, the final and most deadly lethal form of sepsis, with each additional affected organ the mortality increases drastically.

## 2.2. Sepsis-3 Definition

As illustrated in the previous section, it is difficult to pin-point the exact moment at which the immune response switches from normal to dysregulated behavior. To classify patients at septic multiple clinically grounded definitions have been proposed over time. The most up to date and widely used definition, called Sepsis-3, will be introduced in this section. The Sepsis-3 definition as been created by a working group led by the “Society of Critical Care Medicine and the European Society of Intensive Care Medicine” in 2016. Their resolution, named “Third International Consensus Definitions for Sepsis and Septic Shock” [4], hence the name Sepsis-3, provides until today the most widely used sepsis definition and guidance on sepsis identification.

In general, sepsis does not classify as a specific illness, rather a multifaceted condition of “physiologic, pathologic, and biochemical abnormalities” [4], and septic patients are largely heterogeneous. Also the trigger is explicitly nonspecific, since different triggers can cause the same septic condition. Most commonly the underlying cause of sepsis is diarrhoeal disease, the most common underlying injury stems from road traffic injuries and maternal disorders the most common non-communicable disease causing sepsis [1].

According to the Sepsis-3 definition, a patient is in a septic condition if the following two criteria are fulfilled:

1. **Confirmed or Suspected Infection**, which has no strict medical definition or classification, meaning what counts as Suspected Infection (SI) remains a vague. Ultimately it is left for the medical personnel to classify infections or the suspicion of infections. For retrospective data-driven classification it is suggested to characterize any patient prescribed with Antibiotics (ABX) followed by the cultivation of body fluids, or the other way around, with a SI [4].

The timings of prescription and fluid samplings play a crucial role. If the antibiotics were administered first, then the cultivation has to be done in the first 24 hours after first prescription, if the cultivation happened first, the ABX have to be prescribed in the following 72 hours [4]. These timings can be seen in the lower part of Figure 2.2., with the abbreviated ABX. Regardless which happened first, the earlier of the two times is treated as the time of suspected infection onset time.

2. **Dysregulated Host Response** is characterized by the worsening of organ functionality over time. To measure the “amount of dysregulation” the Sepsis-3 consensus relies on the Sequential Organ Failure Assessment (SOFA)-score introduced in [4], [15]. The score is now regularly used to evaluate the functionality of organ systems and helps to predict the risk of mortality, also outside of a sepsis context. The SOFA score is calculated at least every 24 hours and assess six different organ systems by assigning a score from 0 (normal function) to 4 (high degree of dysfunction) to each. The overall score is calculated as sum of each individual system.

It includes the respiratory system, the coagulation/clotting of blood, i.e. changing from liquid to gel, the liver system, the cardiovascular system, the central nervous system and the renal system/kidney function. A more detailed listing of corresponding markers for each organ assessment can be found in table Table 1.9. in the Appendix A.1. The magnitude of a patients initial SOFA-score captures preexisting organ dysfunction. An increase in SOFA score  $\geq 2$  corresponds to an acute worsening of organ functionalities and a drastic worsening in the patients condition, the indicator for a dysregulated response.

The combination of the two criteria represents an exaggerated immune reaction that results in organ dysfunction, when infection is first suspected, even modest organ dysfunction is linked to a 10% increase of in-hospital mortality.

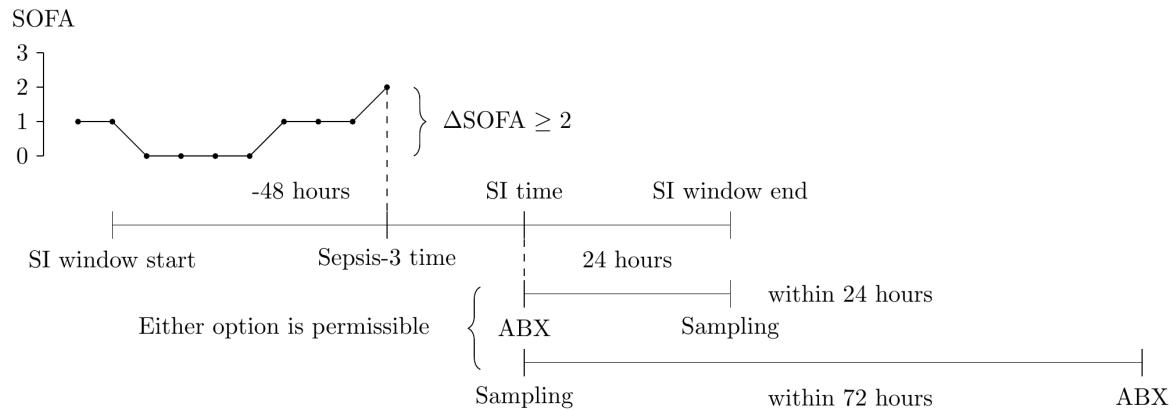
### 2.2.1. Sepsis Classification

The Sepsis-3 definition not only provides the clinical criteria of septic conditions, but also introduces the necessary time windows for sepsis classification. An increase of SOFA  $\geq 2$  in the 48 hours before or 24 hours after the SI time, the so called SI-window, is per Sepsis-3 definition the *sepsis onset time*. A schematic of the timings is shown in figure Figure 2.2..

With respect to which value the increase in SOFA is measured, i.e. the baseline score, is not clearly stated in the consensus and leaves room for interpretation, commonly used approaches include:

- the minimal value inside the SI-window before the SOFA increase,
- the first value of the SI-window,
- the lowest value of the 24h previous to the increase.

Differences in definitions greatly influence the detection of sepsis, which are used for prevalence estimates for example [16]. Using the lowest SOFA score as baseline, the increase  $\geq 2$  for patients with suspected infection was associated with an 18% higher mortality rate according to [15] a retrospective Intensive Care Unit (ICU)-data analysis.



**Figure 2.2.:** Graphical representation of the timings in the Sepsis-3 definition, taken from [17]

Up until today, even though SOFA was created as a clinical bedside score, some of the markers used in it are not always available to measure or at least not at every 24h [18]. For a faster bedside assessment [15] also introduced a clinical score termed Quick Sequential Organ Failure Assessment (qSOFA), with highly reduced marker number and complexity, it includes:

- Respiratory rate  $\geq 22/\text{min}$
- Altered mentation
- Systolic blood pressure  $\leq 100 \text{ mmHg}$

Patients fulfilling at least two of these criteria have an increased risk of organ failure. While the qSOFA has a significantly reduced complexity and is faster to assess it is not as accurate as the SOFA score, meaning it has less predictive validity for in-hospital mortality [15].

There is also the notion of a septic shock, also defined in [4], which an in-hospital mortality above 40%. Patients with a septic shock are can be identified by:

- Sepsis
- Persisting hypotension requiring vasopressors to maintain MAP  $\geq 65\text{mmHg}$
- Serum lactate level  $> 2 \text{ mmol/L}$ , despite volume resuscitation.

## 2.3. Sepsis Prediction

To this day sepsis, and the more extreme septic shock, remains an extreme burden to the worldwide healthcare system. It is associated with high rates of incidence, high mortality and significant morbidity. Despite overall advancements in medical care and slowly decreasing prevalence numbers, sepsis continues to be the leading cause of in-hospital death [19].

In Germany it was estimated in 2022 that at least 17.9% of intensive care patients develop sepsis, and 41.7% of all hospital treated sepsis patients die during their stay [3]. The economic

burden is equally severe, with the annual cost of sepsis treatment in Germany estimated to be €7.7 billion based on extrapolated data from 2013 [3].

Globally sepsis remains to be under-diagnosed significantly due to its nonspecific symptoms. Environmental and socioeconomic factors, such as insufficient sanitation, limited access to clean water and healthcare increase the incidence particularly in low- to middle income countries [1], [19].

Traditional sepsis screening relies on clinical scoring systems such as SOFA or qSOFA. While useful for standardizing assessment, these scores are inherently reactive, since they identify patients already experiencing organ dysfunction rather than those at risk of developing sepsis. This clinical reality has motivated the development of automated prediction systems that can continuously monitor patients and alert clinicians to elevated sepsis risk before overt organ failure develops. Usually, these alerts are based on predetermined criteria which are derived from clinical and laboratory measurements. With the help of these automated alerts, clinicians can potentially more rapidly initiate antibiotic or other treatment or intensify patient monitoring.

### 2.3.1. Limitations of Current Prediction Systems

A meta-analysis of seven sepsis alert systems found no evidence for improvement in patient outcomes, suggesting insufficient predictive power of analyzed alert systems or inadequate system integration [20]. Nevertheless, positive treatment outcomes depend heavily on timely recognition and intervention [19]. Each hour of delayed treatment increases mortality risk, underscoring the critical importance of early detection [21] while structured screening and early warning systems have demonstrated reductions in time-to-antibiotics and improvements in outcomes [22]. These findings confirm that earlier identification of sepsis improves clinical results, even if existing tools are not yet capable enough, and emphasize the need for developing more reliable alert systems.

A recent study suggests a paradigm shift in sepsis detection, namely from a purely symptom-based to a systems-based approach [23]. Instead of waiting for clinical signs, i.e. symptoms, early recognition should integrate multiple physiological and biochemical signals to capture the transition from infection to organ dysfunction. This aligns with the findings of a survey among clinicians regarding Artificial-Intelligence-Assistance in healthcare [24]. According to the survey, one participant emphasizes that specific vitals signs might be of less importance, rather the trend of a patient's trajectory itself should be the prediction target. Another finding of the same study was the preference of trajectories over plain binary event predictions.

Despite these promising directions, the translation of predictive models into clinical practice has proven challenging. Studies consistently identify barriers, such as alert fatigue, where excessive false positives or clinically non-actionable alarms disrupt workflow, reduce clinician trust, and ultimately lead to ignored warnings. Prediction systems face a fundamental trade-off: higher sensitivity captures more true cases but generates more false alarms, while higher specificity reduces alert fatigue but risks missing sepsis cases where early treatment is

critical. To be effective, predictive systems must integrate seamlessly into existing workflows provide interpretable output and support clinical expertise [24].

Taken together, these insights highlight both the need and the opportunity for improved sepsis prediction. The global burden and clinical urgency justify the development of more reliable prediction systems. At the same time, the limitations of current alert systems and implementation barriers underline the necessity for models that can integrate dynamic patient data and capture clinical trajectories. The following chapter examines existing approaches to sepsis prediction and identifies key gaps that motivate the present work.

# 3

## State of the Art

As the last chapter concluded, sepsis represents a critical challenge in modern healthcare, it is both common and deadly, yet hard to diagnose. This chapter provides a brief overview of the current approaches of automated sepsis prediction and the fundamental challenges when comparing these.

Sepsis prediction models for individual patients can be categorized into two major classes, the model-based and the data-driven approaches, each with their own distinct strengths and limitations.

### 3.1. Model-Based Approaches

Biologically and medically inspired models of sepsis offer high interpretability and explainability, since they explicitly encode causal relationships. However, due to the inherent complexity of sepsis pathophysiology, such mechanistic models remain rare [25].

Most existing works focus on dynamic immune response on a cellular level [26], [27], [28], [29]. Complicated signaling and production mechanisms influenced by varying cell concentration are typically modeled using large systems of coupled differential equations.

To derive risk estimates or disease trajectories, model parameters are fitted to individual patient observations. By simulating physiological trajectories under hypothetical infection scenarios, these models enable to estimate the likelihood of sepsis development [27]. More advanced digital twins which incorporate bidirectional feedback between the mechanistic model and patient data have also been explored in [28].

Mechanistic sepsis models are usually validated by comparing simulation trajectories to repetitively measured cellular concentrations. Since the required high-resolution immunological measurements are difficult and costly to obtain, only small patient samples have been validated. To date, no model-based prediction approach has been evaluated on large-scale clinical datasets, limiting insight into real-world performance and generalizability.

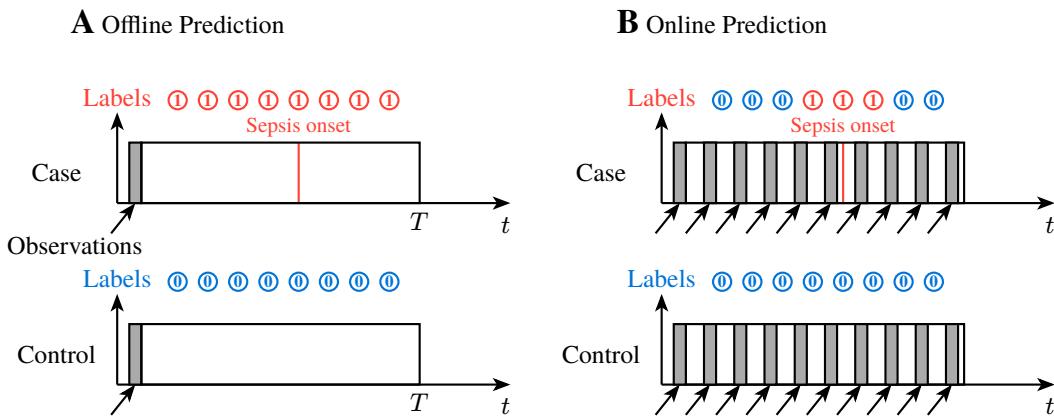
### 3.2. Data-Driven Approaches

With the increasing availability of Electronic Health Record and computational resources, Machine Learning (ML)- and Deep Learning (DL)-methods have become the dominant paradigm for sepsis prediction systems over the last decade. Data-driven approaches can capture

highly nonlinear relationships in heterogeneous clinical data. Unlike mechanistic models, these methods do not require explicit specification of biological relationships, instead, they learn predictive patterns directly from historical data.

Research on data-driven sepsis prediction systems is highly active, in the past five years alone (2021–2026), six systematic reviews on data-driven sepsis predictions have been published [30], [31], [32], [33], [34], [35]. The reviews include a total of 180 studies (including 7 to 73 works per review), proposing over 50 distinct ML and DL methodologies that range from classical to highly specialized methods. The following overview is based primarily on findings from [30], [31], [32] because, taken together, they provide a very comprehensive and complementary coverage of data-driven sepsis prediction research within the considered time frame. By focusing on these three reviews, the analysis captures the majority of relevant studies and conclusions while avoiding redundancy.

Studies differ fundamentally in how they frame the prediction problem, most prominently in *online prediction* versus *offline prediction*. In online prediction newly arriving medical measurements are incorporated into a continuously updated risk estimate., In offline prediction only the information available at a fixed observation time is used to predict the risk of sepsis within a pre specified time-horizon  $T$ . Because these setups rely on different information structures and temporal assumptions, their reported performances are not directly comparable. Online prediction is more clinically relevant but also more challenging to implement given the current state of available Electronic Health Record (EHR). Both schemes are shown in Figure 3.3., note that in offline prediction the horizon  $T$ , the specific choice strongly influences the outcome, with smaller horizons the tasks becomes gradually easier. For the online scheme choice of what time range around a diagnosed sepsis onset qualifies as positive label influences prediction accuracy.



**Figure 3.3.:** Illustration of the two predictions schemes, *offline* (A) vs. *online* (B) (figure heavily inspired by [31]). The main difference is the sepsis labeling, as well as provision and utilization and arrival of observation data.

Most sepsis prediction models are trained in a retrospectively and evaluated using offline prediction tasks, typically predicting sepsis onset  $T = 6\text{--}48$  hours in advance. Model performance is commonly reported using Area Under Receiver Operationg Curve (AUROC)

and Area Under Precision Recall Curve (AUPRC) (the metric derivation and interpretation is discussed in Section 5.5). Across studies, reported AUROC values typically range from approximately 0.60 to 0.95, indicating modest to very good performance, though such values must be interpreted cautiously given differences in cohort definition, task formulation and evaluation protocols. To give a comparison, classical assessment scores achieve AUROCs of SOFA= 0.667 and qSOFA= 0.612 [32].

Methodologically, a wide range of supervised learning approaches has been explored. Classical models such as logistic regression, Cox proportional hazards models, and random forests or gradient boosting remain strong baselines due to their robustness and interpretability. Deep learning architectures, including Recurrent Neural Networks (RNN), temporal convolutional networks, and more recently transformer-based models, have been proposed to capture complex temporal dependencies. In general, explainability of these predictions predominantly rely on Shapley-values analyses, deriving importance factors of single input-features or input-feature interactions [35], [36].

Most models rely on routinely collected clinical data, including vital signs, laboratory measurements, demographics, and treatment variables. Publicly available ICU datasets, for example the Medical Information Mart for Intensive Care (MIMIC) [37] series, serve as the predominant development and benchmarking platforms. Differences in feature selection substantially influences both model performance and real-world usability. While a broader set of features can increase predictive accuracy but risks again the clinical applicability, if required measurements are not routinely available. Moreover, extensive feature sets increase the risk of label leakage, where the measurements and medical concepts used to derive the sepsis label are provided to the prediction model as feature input. This way the model would learn the sepsis derivation but not underlying signals which are actually helpful for early sepsis recognition. As feature selection is not standardized, the reviewed works deployed feature sets of size 2 to 100, again emphasizing the heterogeneous nature field of research.

Finally, due to ambiguities in the Sepsis-3 definition, the deployed definitions vary widely across studies and greatly influence prevalence, cohort composition and therefore the task difficulty. Intuitively, different sepsis definitions are not comparable with each other since they might capture dissimilar medical concepts. Even for the same conceptual definition and same dataset differences in implementation can yield different patient cohorts and therefore different prediction performances [16]. More restrictive definitions typically produce lower prevalence and greater class imbalances making ML-based prediction more difficult but potentially increasing clinical relevance. Less restrictive definitions can artificially inflate prediction performance while reducing practical applicability. Since less than 10% of works publish their code for label generation, hurting reproducibility and making comparisons often impossible [31].

Overall the field of research on data-driven sepsis prediction is highly relevant and active. To date, it has generated numerous heterogeneous methodologies, where most of these works provide proof-of-concepts, and only the smallest fraction is implemented into clinical practice for further investigation. A major challenge remains the incomparability and lack of standardization in model development and evaluation.

### 3.3. Summary of State of the Art

As discussed, purely model-based and purely data-driven come with their own sets of strengths and limitations. Mechanistic models offer strong interpretability and encode physiological priors, yet their practical usefulness is limited by the scarcity of high-resolution immunological measurements and the lack of large-scale clinical validation. In contrast, data-driven models show strong empirical performance on EHR datasets, but their prediction behavior is often difficult to interpret and can show black-box behavior.

This work, aims to combine the strengths of both paradigms: mechanistic components of a physiologically inspired model is used to introduce structured physiological biases to help the learning process and provide more interpretable intermediate quantities. At the same time, the data-driven components allow the model to adapt to real clinical variability and make use of information that is not explicitly captured by the mechanistic structure. In this way, this work novel methodology seeks to make data-driven sepsis prediction models more transparent and more robust.

The heterogeneity in prediction tasks, sepsis definitions and feature sets illustrates why each decision of one of these aspects regarding new prediction models should be taken with care. For the sake of reproducibility the choices need to be reported as precisely as possible. Works such as Yet Another ICU Benchmark (YAIB) [38] attempt to address these challenges by providing a common basis for evaluating models by standardizing the dataset, cohort definition, prediction task, and labeling strategy, thereby enabling fair and reproducible comparison of different approaches.

After introducing the methodology of this work in Section 5, its performance validated on real clinical data in Section 6. The experimental setup makes use of the YAIB framework, and its dataset, cohort definition, prediction task, and labeling implementation are described in detail in Section 6.1, as these settings are adopted for training and evaluating the developed method.

# 4

## Model Background (Dynamic Network Model)

As outlined in Section 2, the macroscopic multi-organ failure associated with sepsis is driven by a dysregulated cascade of signaling processes on a microscopic level (see Section 2.1). This cascade involves a massive amount of interconnected components, where the connections mechanics and strengths vary over time and space. For example, these interactions differ across tissues and evolve as sepsis progresses, with crossing biochemical thresholds the behavior of cells can be changed [39].

In essence, cell-to-cell and cell-to-organ interaction in septic conditions form a highly dynamic, nonlinear and spatio-temporal network of relationships [25], which cannot be fully understood by a reduction to single time-point analyzes. Even though many individual elements of the inflammatory response are well characterized, we still fail to integrate them into a coherent system-level picture.

To address this complexity, the emerging field of *Network Physiology* provides a promising conceptual framework. Rather than studying components in isolation, network physiology focuses on the coordination and interconnection among the diverse organ systems and subsystems [40]. It enables the study of human physiology as a complex, integrated system, where emergent macroscopic dynamics arise from interacting subsystems that cannot be explained by their individual behavior. This perspective translates to the mesoscopic level, i.e. the in-between of things, where the coupling mechanisms collectively determine the overall physiological function.

In network physiology, the analytical basis of the bodies interacting systems is often graph based. Nodes represent subsystem such as organs or cell populations and links represent functional couplings or communication pathways [40]. Unlike classical graph theory, where dynamics are introduced by changing the graph topology (e.g. adding or removing links or nodes), in *Complex Networks* the links themselves can evolve dynamically in response to other system variables. These adaptive connections allow for information to propagate through the whole network, giving rise to emerging phenomena on global scales for otherwise identical network topologies.

Complex networks are well studied in physics and biology and have been applied to various physiological domains. Early works, such as [41] that have studied the cardiovascular system, while more recent studies have focused on the cardio-respiratory coupling [42] and large-scale brain network dynamics [43]. Network approaches have also provided mechanistic insights into disease dynamics, for example Parkinson [44] and Epilepsy [45].

Building on these interaction centric principles has opened up new opportunities to study how the inflammatory processes, such as those underlying sepsis, emerge from the complex inter- and intra-organ communication. In particular Sawicki et. al [46] and Berner et. al [47] have introduced a dynamical system that models the cytokine behavior in patients with sepsis and cancer. This functional model will be referred to as Dynamic Network Model and forms the conceptual foundation for this whole project.

The remainder of this chapter is structured as follows: In Section 4.1 introduces the theoretical backbone of the Dynamic Network Model (DNM), the Kuramoto oscillator model, which provides a minimal description of synchronization phenomena in complex systems. Section 4.2 presents the formal mathematical definition of the DNM and its medical interpretation, followed by implementation details in Section 4.3 and a presentation of selected simulation results in Section 4.3.4.

## 4.1. Theoretical Background: The Kuramoto Oscillator Model

To mathematically describe natural or technological phenomena, *coupled oscillators* have proven to be a useful framework, for example, to model the relative timing of neural spiking, reaction rates of chemical systems or dynamics of epidemics [48]. In these cases complex networks of coupled oscillators are often capable of bridging microscopic dynamics and macroscopic synchronization phenomena observed in biological systems.

One of the most influential system of coupled oscillators is the *Kuramoto Phase Oscillator Model*. This system of coupled Ordinary Differential Equation (ODE), where each term describes an oscillator rotating around the unit circle, is often used to study how synchronization emerges from simple coupling rules. In the simplest form, it consists of  $N \in \mathbb{N}_{>0}$  identical, fully connected and coupled oscillators with phase  $\varphi_i \in [0, 2\pi)$ , for  $i \in 1 \dots N$  and an intrinsic frequency  $\omega_i \in \mathbb{R}$  [49]. The phase  $\varphi_i$  represents the angular position on the unit circle, hence it is defined modulo  $2\pi$ . The dynamics are given by the ODE system:

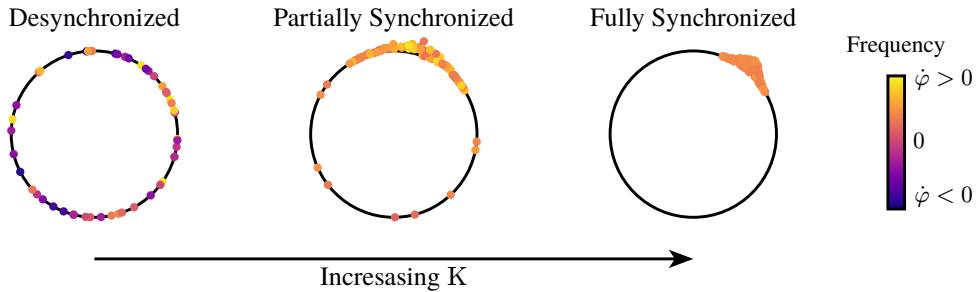
$$\dot{\varphi}_i = \omega_i - \frac{K}{N} \sum_{j=1}^N \sin(\varphi_i - \varphi_j) \quad (4.1)$$

Here, the  $\dot{\varphi}$  is used as shorthand notation for the time derivative of the phase  $\frac{d\varphi}{dt}$ , the instantaneous phase velocity. An additional parameter is the global coupling strength  $K \in \mathbb{R}_{>0}$  between oscillators  $i$  and  $j \in 1 \dots N$ . The system evolution depends on the choice of initial phases  $\varphi_i(t=0)$ , which are typically drawn from a uniform random distribution over  $[0, 2\pi)$ .

When evolving this system with time, commonly by numerical integration, since analytical solutions only exist for special cases, oscillator  $i$ 's phase velocity depends on each other oscillator  $j$ . The sine coupling  $\sin(\varphi_i - \varphi_j)$  implements a phase-cohesion mechanism: oscillator  $i$  decelerates when it leads oscillator  $j$  ( $0 < \varphi_i - \varphi_j < \pi$ ) and accelerates when it lags behind ( $-\pi < \varphi_i - \varphi_j < 0$ ), pulling the population towards synchronization.

For sufficiently large  $N$  the oscillator population can converge towards system-scale states of coherence or incoherence based on the choice of  $K$  [50]. Coherent in this case means

oscillators synchronize with each other, so they share the same phase and phase velocity, incoherence on the other hand is the absence of synchronization (desynchronized), see Figure 4.4.. Synchronous states can be reached if the coupling is stronger than a certain threshold  $K > K_c$ , the critical coupling strength [50]. In between these two regimes there is a transition-phase of partial synchronization, where some oscillators phase- and frequency-lock and others do not. The model captures the mechanism of self-synchronization, and a collective transition from disorder to order, that underlies many real world processes, which is the reason the model has attracted so much research [48].



**Figure 4.4.:** Illustration of the collective dynamics of a population of phase oscillators on the unit circle as the global coupling strength  $K$  is increased. Each point represents an oscillator at phase  $\varphi_i$ , color encodes the instantaneous phase velocity  $\dot{\varphi}_i$  (blue/yellow: faster, green: slower). For weak coupling (left), oscillators are desynchronized, with phases distributed around the circle and heterogeneous frequencies. At intermediate coupling (center), partial synchronization emerges: a subset of oscillators forms a coherent cluster that phase- and frequency-locks, while the remaining oscillators drift incoherently. For sufficiently strong coupling (right), the population becomes fully synchronized, with all oscillators sharing a common phase and frequency.

#### 4.1.1. Extensions to the Kuramoto Model

To increase the generalization ability of the system, various extensions of the basic Kuramoto model have been proposed and studied numerically and analytically. Several extensions are directly relevant to the DNM and their definitions and effects on synchronization will be introduced, with additional terms being indicated by the red color:

**Phase Lag  $\alpha$**  introduced in [51], where  $|\alpha| < \frac{\pi}{2}$ . By shifting the coupling function, values of  $\alpha \neq 0$  act as an inhibitor of synchronization, so the coupling does not vanish even when the phases align:

$$\dot{\varphi}_i = \omega_i - \frac{K}{N} \sum_{j=1}^N \sin(\varphi_i - \varphi_j + \color{red}{\alpha}) \quad (4.2)$$

As a result the critical coupling strength  $K_c$  increases with  $\alpha$  [51].

**Adaptive coupling**  $\mathbf{K} \in [-1, 1]^{N \times N}$  moves from a global coupling strength  $K$  for all oscillator pairs to an adaptive asymmetric coupling strength for each individual pair  $\mathbf{K} = (\kappa_{ij})_{(i,j) \in N \times N}$ :

newpage?

$$\dot{\varphi}_i = \omega_i - \frac{1}{N} \sum_{j=1}^N \kappa_{ij} \sin(\varphi_i - \varphi_j) \quad (4.3)$$

$$\dot{\kappa}_{ij} = -\varepsilon (\kappa_{ij} + \sin(\varphi_i - \varphi_j + \beta^\mu)) \quad (4.3)$$

Here adaption rate  $0 < \varepsilon \ll 1$  separates the fast moving oscillator dynamics from slower moving coupling adaptivity [52]. Such adaptive couplings have been used to model neural plasticity and learning-like processes in physiological systems [53]. The so called new phase lag parameter  $\beta$  of the adaptation function (also called plasticity rule) plays an essential role in the synchronization process. At a value of  $\beta^\mu = \frac{\pi}{2}$  the coupling, and therefore the adaptivity, is at a maximum positive feedback, strengthening the link  $\kappa_{ij}$  and encouraging synchronization between oscillators  $i$  and  $j$ . This maximal connectivity is referred to *Hebbian Rule* and found in synchronizing systems such as the brain [52]. For other values  $\beta^\mu \neq \frac{\pi}{2}$  the feedback is delayed  $\varphi_i^\mu - \varphi_j^\nu = \beta^\mu - \frac{\pi}{2}$  by a phase lag, a value of  $\beta^\mu = -\frac{\pi}{2}$  we get an anti-Hebbian rule which inhibits synchronization.

**Multiplex Networks** represent systems with multiple interacting layers. Multiplexing introduces a way to couple  $L \in \mathbb{N}_{>1}$  Kuramoto networks (indexed by  $\mu = 1, \dots, L$  and  $\nu = 1, \dots, L \neq \mu$ ) via interlayer links:

$$\dot{\varphi}_i^\mu = \omega_i - \frac{K}{N} \sum_{j=1}^N \sin(\varphi_i - \varphi_j + \alpha^\mu) - \sigma^{\mu\nu} \sum_{\nu=1, \nu \neq \mu}^L \sin(\varphi_i^\mu - \varphi_i^\nu + \alpha^{\mu\nu}) \quad (4.4)$$

Here  $\mu$  and  $\nu$  represent distinct subsystems, and are connected via interlayer coupling weights  $\sigma^{\mu\nu} \in \mathbb{R}_{\geq 0}$ , where oscillators are acting one-to-one.

These extensions combined serve as the source of dynamics for the DNM and give rise to more intricate system states than the straightforward synchronization in the base model. Even for single layers, non-multiplexed but phase-lagged and adaptively coupled oscillators, one can observe several distinct system states, neither fully synchronized or desynchronized such as phase and frequency-clusters, chimera- and splay states. The emergence of these states depends on the choice of the coupling strength  $K$  and the phase-lag parameters  $\alpha$  and  $\beta$ .

In the frequency clustered state, the oscillator phases do not synchronize, but several oscillator groups can form that share a common frequency. For the phase-clustered case, the groups additionally synchronize their phase. Frequency clusters often emerge as intermediate regimes between full synchronization and incoherence [54].

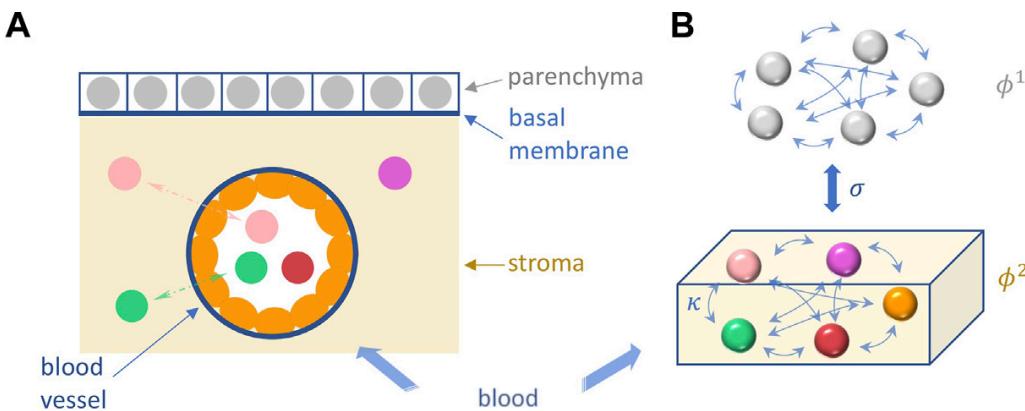
Chimera states, a special type of partial synchronization, occur when only a subset of oscillators synchronizes in phase and frequency, while others remain desynchronized. In contrast to “normal” partial synchronization they occur when the coupling symmetry breaks. In splay states, all oscillators synchronize their frequencies but do not their phases, they instead uniformly distribute around the unit circle [52].

The introduction changes the system behavior once more, for example single layers of a multiplexed system can result in the multi-clustered regime for parameters they would not in the monoplexed case. In multiplexed systems it is also possible connected layers end up in different stable state, for example, one in a clustered the other in a splay state.

## 4.2. Description

The DNM is a **functional** model, that means it **does not try to model things accurately on any cellular, biochemical, or organ level**, it instead tries to model dynamic interactions. At the core, the model does differentiate between two broad classes of cells, introduced in Section 2.1.1, the stroma and the parenchymal cells. It also includes the cell interaction through cytokine proteins and an information flow through the basal membrane. Importantly, the model only handles the case of already infected subjects and tries to grasp if the patients state is prone to a dysregulated host response.

Cells of one type are aggregated into layers, everything associated with parenchymal cells is indicated with an  $.^1$  superscript and is called the *organ layer*, stroma cells are indicated with  $.^2$  and is referred to as nonspecific *immune layer*. Each layer consists of  $N$  phase oscillators  $\varphi_i^{1,2} \in [0, 2\pi)$ . To emphasize again the function aspect of the model: individual oscillators do not correspond to single cells, rather the layer as a whole is associated with the overall state of all organs or immune system functionality respectively. An illustration of the biological and functional model is shown in Figure 4.5..



**Figure 4.5.:** Image taken from [47] in a simplified form, illustrating the Dynamic Network Model. (A) A tissue element is depicted, in which the basic processes of sepsis take place. Shown are different immune and structural cells involved (colored) in the stroma (yellow area), the parenchyma (grey), and the capillary blood vessel, where cytokines, PAMPs and DAMPs are transmitted. (B) Depicts the functional interactions within and between the two corresponding network layers in the DNM, the parenchyma and the stroma (immune layer).

The metabolic cell activity is modeled by rotational velocity  $\dot{\varphi}$  of the oscillators, the faster the rotation, the faster the metabolism. Each layer is fully coupled via an adaptive possibly asymmetric matrix  $\mathbf{K}^{1,2} \in [-1, 1]^{N \times N}$  with elements  $\kappa_{ij}^{1,2}$ , these couplings represent the activity of cytokine mediation. Small absolute coupling values indicate a low communication via cytokines and grows with larger coupling strength. For the organ layer there is an additional non-adaptive coupling part  $\mathbf{A}^1 \in [0, 1]^{N \times N}$  with elements  $a_{ij}^1$ , representing a fixed connectivity within an organ.

The dimensionless system dynamics are described with the following coupled ODE terms, build on the classical Kuramoto model described in Section 4.1 and its extensions from Section 4.1.1:

$$\dot{\varphi}_i^1 = \omega^1 - \frac{1}{N} \sum_{j=1}^N \{ (a_{ij}^1 + \kappa_{ij}^1) \sin(\varphi_i^1 - \varphi_j^1 + \alpha^{11}) \} - \sigma \sin(\varphi_i^1 - \varphi_i^2 + \alpha^{12}) \quad (4.5)$$

$$\dot{\kappa}_{ij}^1 = -\varepsilon^1 (\kappa_{ij}^1 + \sin(\varphi_i^1 - \varphi_j^1 - \beta)) \quad (4.6)$$

$$\dot{\varphi}_i^2 = \omega^2 - \frac{1}{N} \sum_{j=1}^N \{ \kappa_{ij}^2 \sin(\varphi_i^2 - \varphi_j^2 + \alpha^{22}) \} - \sigma \sin(\varphi_i^2 - \varphi_i^1 + \alpha^{21}) \quad (4.7)$$

$$\dot{\kappa}_{ij}^2 = -\varepsilon^2 (\kappa_{ij}^2 + \sin(\varphi_i^2 - \varphi_j^2 - \beta)) \quad (4.8)$$

Where the interlayer coupling, i.e. a symmetric information through the basal lamina, is modeled by the parameter  $\sigma \in \mathbb{R}_{\geq 0}$ . The internal oscillator frequencies are modeled by the parameters  $\omega^{1,2}$  and correspond to a natural metabolic activity.

Besides the coupling weights in  $\mathbf{K}^{1,2}$  the intralayer interactions also depend on the phase lag parameters  $\alpha^{11}$  and  $\alpha^{22}$  modeling cellular reaction delay. To separate the fast moving oscillator dynamics from the slower moving coupling weights adaption rates  $0 < \varepsilon \ll 1$  are introduced. Since the adaption of parenchymal cytokine communication is assumed to be slower than the immune counterpart [46], it is chosen  $\varepsilon^1 \ll \varepsilon^2 \ll 1$ , which introduces dynamics on multiple timescales.

Lastly, the most influential parameter is  $\beta$  which controls they adaptivity of the cytokines. Because  $\beta$  has such a big influence on the model dynamics it is called the (*biological*) *age parameter* and summarizes multiple physiological concepts such as age, inflammatory baselines, adiposity, pre-existing illness, physical inactivity, nutritional influences and other common risk factors [47].

All the systems variables and parameters are summarized in Table 4.1. together with their medical interpretation.

**Table 4.1.:** Summarization of notation used in the Dynamic Network Model. Superscripts indicating the layer are left out for readability.

SYMBOL	NAME	PHYSIOLOGICAL MEANING
<b>Variables</b>		
$\varphi_i$	Phase	Group of cells
$\dot{\varphi}_i$	Phase Velocity	Metabolic activity
$\kappa_{ij}$	Coupling Weight	Cytokine activity
<b>Parameters</b>		
$\alpha$	Phase lag	Metabolic interaction delay
$\beta$	Plasticity rule	Combined of risk factors
$\omega$	Natural frequency	Natural cellular metabolism
$\varepsilon$	Time scale ratios	Temporal scale of cytokine activity

SYMBOL	NAME	PHYSIOLOGICAL MEANING
$a_{ij}$	Connectivity	Fixed intra-organ cell-to-cell interaction
$\sigma$	Interlayer coupling	Interaction between parenchymal and immune cells through the basal lamina
<b>Measures</b>		
$s$	Standard deviation of frequency (see Equation (4.15))	Pathogenicity (Parenchymal Layer)

#### 4.2.1. Pathology in the DNM

A biological organism, such as the human body, can be regarded as a self-regulating system that, under healthy conditions, maintains a homeostatic state [6]. Homeostasis refers to a dynamic but balanced equilibrium in which the physiological subsystems continuously interact to sustain stability despite external perturbations. In the context of the DNM, this equilibrium is represented by a synchronous regime of both layers in the duplex oscillator system. In synchronous states, the organ layer and immune layer exhibit coordinated phase and frequency dynamics, reflecting balanced communication, collective frequency of cellular metabolism and stable systemic function.

Pathology, in contrast, is modeled by the breakdown of the synchronicity and the formation of frequency clusters in the parenchymal layer, i.e. loss of homeostatic balance. In the DNM least one cluster will exhibit increased frequency and one with lower or unchanged frequency. This aligns with medical observation, where unhealthy parenchymal cells change to a less efficient anaerobic glycosis based metabolism, forcing them to increase their metabolic activity to keep up with the energy demand. Remaining healthy cells are expected to stay frequency synchronized to a lower and “healthy” frequency.

There are two more cases, neither fully healthy nor fully pathologic, representing a vulnerable or resilient patient condition. The healthy but vulnerable case corresponds to a splay state, where phases in the parenchymal layer are not synchronized, but the frequencies are, weakening the overall coherence [47]. A resilient state corresponds to cases where both the phase and frequency of the parenchymal layer are synchronized, but the immune layer exhibits both frequency and phase clustering.

It is important to note, that the ODE dynamics or system variable trajectories **do not** translate to the evolution of a patients pathological state. Instead, the amount of desynchronization of the parenchymal layer when reaching a steady system state can be interpreted as the current state of a patients organ functionality. The “result” or solution of the coupled oscillator system does not provide any temporal insights, but only describe the current condition. Time-steps taken inside the model cannot be compared to any real-world time quantity.

## 4.3. Implementation

This subsection describes the implementation for the numerical integration of the DNM defined in Equation (4.8), the choice of initial parameter values and how (de-)synchronicity/disease severity is quantified. One goal of this implementation is to partly reproduce the numerical results presented in [47], since they will be serving as a basis for following chapters.

### 4.3.1. Implementation Details

The backbone for the present numerical integration is JAX [55], a Python package for high-performance array computation, similar to NumPy or MATLAB but designed for automatic differentiation, vectorization and Just In Time Compilation (JIT). JIT-compilation and vectorization allow high-level numerical code to be translated to highly optimized accelerator-specific machine code, for example Graphics Processing Unit (GPU). This way, performance benefits of massively parallel hardware can be utilized with minimal extra programming cost. For the actual integration a differential equation solver from diffrax [56] was used, which provides multiple solving schemes fully built on top of JAX.

While [47] uses a fourth-order Runge-Kutta method and a fixed step-size, this implementation<sup>1</sup> uses the Tsitouras 5/4 Runge-Kutta method [57] with adaptive step-sizing controlled by a Proportional-Integral-Derivative (PID) controller, allowing for more efficient integration while keeping similar accuracy. A relative tolerance of  $10^{-3}$  and an absolute tolerance  $10^{-6}$  were chosen. All simulations were carried out in 64-bit floating point precision, necessary for accurate and stable system integration.

Because of the element-wise differences used in the coupling terms  $\varphi_i^{1,2} - \varphi_j^{1,2} \in \mathbb{R}^{N \times N}$  the computational cost scales quadratically with the number of oscillators  $N$ . These differences are then transformed by the computationally expensive trigonometric sin routine. To accelerate integration, these trigonometric evaluations were optimized following [58]. Terms in the form  $\sin(\theta_l - \theta_m)$  were expanded as:

$$\sin(\theta_l - \theta_m) = \sin(\theta_l) \cos(\theta_m) - \cos(\theta_l) \sin(\theta_m) \quad \forall l, m \in \{1, \dots, N\} \quad (4.9)$$

By caching the terms  $\sin(\theta_l)$ ,  $\sin(\theta_m)$ ,  $\cos(\theta_l)$ ,  $\cos(\theta_m)$  once per iteration, the number of trigonometric evaluations per iteration is reduced from  $2 \cdot [N(N - 1)]$  to  $2 \cdot [4N]$ , significantly improving performance for mid to large oscillator populations.

Additionally, an alternative implementation based on Lie-algebra formulations was also explored, utilizing their natural representation for rotations in N-D-space. Although theoretically promising in terms of numerical accuracy and integration stability, this approach did not yield practical advantages in performance. Further details on this reformulation are provided in Appendix A.2.

[schreiben](#)

---

<sup>1</sup>The code is available at [https://github.com/unartig/sepsis\\_osc/tree/main/src/sepsis\\_osc/dnm](https://github.com/unartig/sepsis_osc/tree/main/src/sepsis_osc/dnm)

### 4.3.2. Parameterization and Initialization

The DNM is dimensionless and not bound to any physical scale, that means there is no medical ground truth of parameter values and their choice is somewhat arbitrary. For the present implementation the parameterization is adopted from the original works [46] and [47] since they have already shown desired properties of (de-)synchronization and valid medical interpretations for these parameter choices.

The majority of their parameter choices heavily simplify the model. First of all, the different natural frequencies are treated as equal and are set to 0 giving  $\omega^1 = \omega^2 = \omega = 0$ , any other choice of  $\omega$  just changes the frame of reference (co-rotating frame), the dynamics stay unchanged [47]. The phase lag parameters for the inter layer coupling are both set to  $\alpha^{12} = \alpha^{21} = 0$ , yielding instantaneous interactions, the intralayer phase lags are set to  $\alpha^{11} = \alpha^{22} = -0.28\pi$ , which was the prominently used configuration in [47] yielding the desired dynamical properties. The constant intralayer coupling in the parenchymal is chosen as global coupling  $a_{ij} = 1$  if  $i \neq j$  else 0.

The adaptation rates are chosen as  $\varepsilon^1 = 0.03$  and  $\varepsilon^2 = 0.3$ , creating the two dynamical timescales for slow parenchymal and faster immune cells. The number of oscillators per layer is chosen as  $N = 200$  throughout all simulations. To account for the randomly initialized variables, each parameter configuration is integrated for an ensemble of  $M = 50$  initializations. This randomization of initial values is used to account for epistemic uncertainties, i.e. systemic errors introduced by modeling simplifications.

In [47] the influence of parameter values for  $\beta$  and  $\sigma$  was investigated and not constant throughout different simulations, with  $\beta \in [0.4\pi, 0.7\pi]$  and  $\sigma \in [0, 1.5]$ , in this work the interval for  $\beta$  was increased to  $[0.0, 1.0\pi]$ . An exhaustive summary of all variable initializations and parameter choices can be found in Table 4.2..

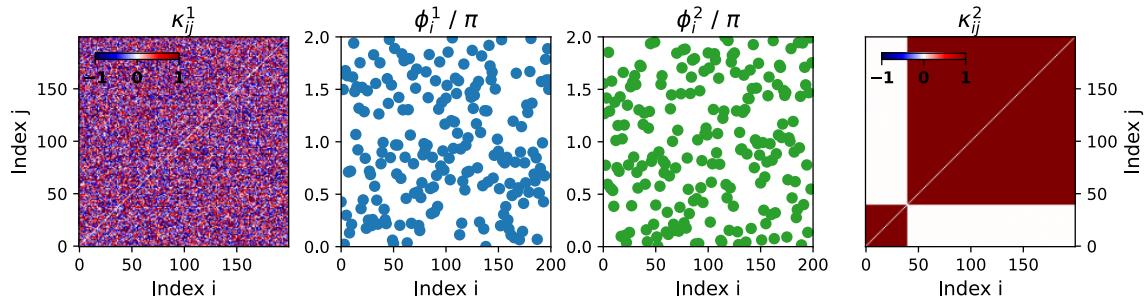
**Table 4.2.:** Parameterization and initialization of the DNM used for the numerical integration.

SYMBOL	VALUE	SYMBOL	VALUE
<b>Variables</b>			
$\varphi_i^1$	$\sim \mathcal{U}[0, 2\pi)$	$\kappa_{i \neq j}^1$	$\sim \mathcal{U}(-1, 1)$
$\varphi_i^2$	$\sim \mathcal{U}[0, 2\pi)$	$\kappa_{i \neq j}^2$	clusters of size $C$ and $1 - C$
<b>Parameters</b>			
$M$	50	$N$	200
$C$	20%		
$\beta$	$[0.0, 1.0]\pi$	$\sigma$	$[0.0, 1.5]$
$\alpha^{11}, \alpha^{22}$	$-0.28\pi$	$\alpha^{12}, \alpha^{21}$	0.0
$\omega_1, \omega_2$	0.0	$A^1$	$\mathbb{1} - I$
$\varepsilon^1$	0.03	$\varepsilon^2$	0.3

Initial values for the system variables, i.e. the phases and coupling strengths, were not parameterized explicitly, rather sampled from probability distributions. The initial phases  $\varphi(0)_i^{1,2}$  are randomly and uniformly distributed around the unit circle for both layers, i.e.  $\varphi(0)_i^{1,2} \sim$

$\mathcal{U}[0, 2\pi]$ . The intralayer coupling of the parenchymal layer coupling is also chosen randomly and uniformly distributed in the interval  $[-1.0, 1.0]$ . Since there is no self-coupling, the diagonal is set to 0.

For the immune layer an initial cytokine activation is modeled by clustering the initial intralayer coupling matrix. A smaller cluster of  $C \cdot N$  oscillators and a bigger cluster of  $(1 - C) \cdot N$  cells. Within the clusters oscillators are connected but not between the clusters. Following [47], the cluster size  $C \in [0, 0.5]$  was chosen as 0.2, but as their findings suggest the size of the clusters does not have impact on the systems dynamics. An example for initial variable values of a system with  $N = 200$  and  $C = 0.2$  is shown in Figure 4.6..



**Figure 4.6.:** Initializations for the variable values of a DNM with  $N = 200$  oscillators per layer. The middle two plots show the phases of the oscillators, with  $\varphi_i^1$  for parenchymal and  $\varphi_i^2$  for the immune layer, sampled from a uniform random distribution from 0 to  $2\pi$ . On the left-hand side is the initialization of the parenchymal intralayer coupling matrix  $\mathbf{K}^1$  from a uniform distribution in the interval from  $-1$  to  $1$ . On the right-hand side is the two cluster initialization for the coupling matrix  $\mathbf{K}^2$  of the immune layer, with a cluster size of  $C = 0.2$ , where each cluster is intra-connected, but without connections between the clusters.

To average out the influence of specific random initial values, simulations are performed for ensembles, combining  $m \in 1, \dots, M$  ensemble members. Throughout this work an ensemble size of  $M = 50$  was used.

### 4.3.3. Synchronicity Metrics

As introduced in Section 4.1, for the complex Kuramoto networks the synchronization behavior is usually the point of interest, in the following two metrics are introduced, relevant to connect the DNM-dynamics to sepsis. There are two relevant states or system configurations that should be identifiable and quantifiable to allow qualified state analyzes: phase and frequency synchronization, for each a distinct measure is required.

**Phase synchronization** of a layer is commonly measured by the *Kuramoto Order Parameter* [50]:

$$R_2^{1,2}(t) = \frac{1}{N} \left| \sum_j^N e^{i \cdot \varphi_j^{1,2}(t)} \right| \quad \text{with } 0 \leq R_2^{1,2} \leq 1 \quad (4.10)$$

where  $R_2^\mu = 0$  corresponds to total desynchronization, the splay-state and  $R_2^\mu = 1$  corresponds to fully synchronized state, for convenience from now on the subscript  $\cdot_2$  is omitted, denoting the Kuramoto Order Parameter simply as  $R^{1,2}$ .

**Frequency synchronization** measurements are more involved, as a starting point first the notion of a layers *mean phase velocity* has to be introduced, which can be calculated as follows:

$$\bar{\omega}^{1,2} = \frac{1}{N} \sum_j^N \dot{\varphi}_j^{1,2}(t) \quad (4.11)$$

The original definition uses an approximated version using the oscillators mean velocity:

$$\langle \widetilde{\dot{\varphi}_j^{1,2}(t)} \rangle = \frac{\varphi_j^{1,2}(t+T) - \varphi_j^{1,2}(t)}{T} \quad (4.12)$$

$$\widetilde{\bar{\omega}^{1,2}(t)} = \frac{1}{N} \sum_j^N \langle \widetilde{\dot{\varphi}_j^{1,2}(t)} \rangle \quad (4.13)$$

for some averaging time window  $T$ . But since their choice of  $T$  is not documented while having substantial influence on the calculation the instantaneous angle velocity from Equation (4.11) was preferred. Due to this change in definition, while the overall systematics stay the same, deviations from their original results are expected.

One can now calculate the *standard deviation of the mean phase velocities*:

$$\sigma_\chi(\bar{\omega}^{1,2}, t) = \sqrt{\frac{1}{N} \sum_j^N (\dot{\varphi}_j^{1,2}(t) - \bar{\omega}^{1,2}(t))^2} \quad (4.14)$$

Where  $\sigma_\chi = 0$  indicates full frequency synchronization and growing values indicate desynchronization and/or clustering. Nonzero values only reveal that there is some desynchronization of the frequencies, but it remains unknown if it is clustered, multi-clustered or fully desynchronized.

Having multiple ensemble members  $m = 1, \dots, M$  with the same parameterization, it is expected that different initialization, meaning initial-values drawn from the parameterized distributions, exhibit dissimilar behaviors, one can also calculate the *ensemble averaged standard deviation of the mean phase velocity*:

$$s^{1,2}(t) = \frac{1}{M} \sum_m^M \sigma_\chi(\bar{\omega}_m^{1,2}, t) \quad (4.15)$$

In [47] it was shown numerically that the quantity  $s^{1,2}$  is proportional to the fraction of ensemble members that exhibit frequency clusters containing at least one oscillator. This makes  $s^1$  a viable measure for pathology, as increasing values of  $s^1$  or increasing system incoherence then indicate more dysregulated host responses and consequently higher risks of multiple organ failure.

#### 4.3.4. Simulation Results

The original findings of [47] identify  $\beta$ , the combined age parameter, and  $\sigma$ , the interlayer coupling strength which models the cytokine activity, as naturally important parameters in order to understand underlying mechanisms of sepsis progression. In the following subsection multiple simulation results are presented, starting with time-snapshots for different parameterization and initializations. Afterward, the transient and temporal behavior of the metrics  $s^{1,2}$  and  $R^{1,2}$  is for the same parameterization, as well as the introduction of the  $\beta, \sigma$  phase space of these metrics.

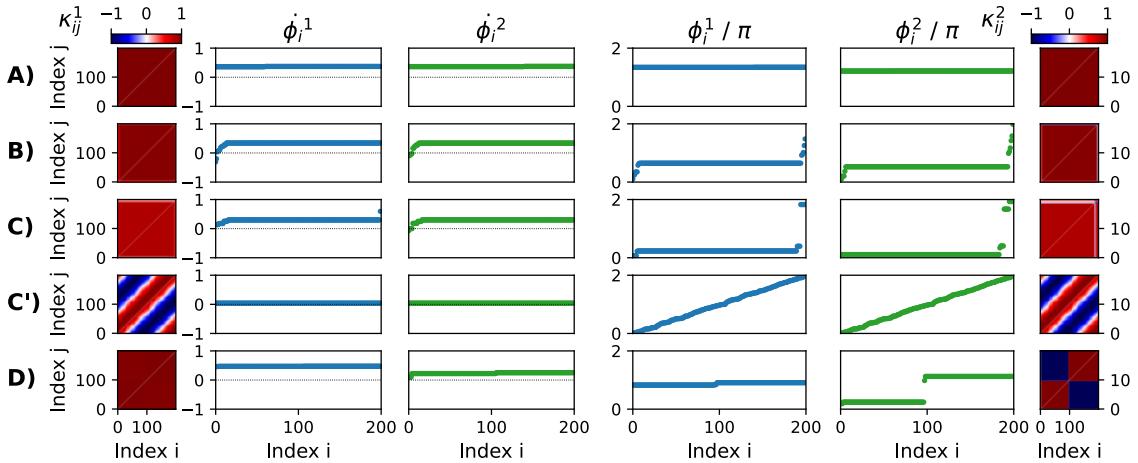
In Figure 4.7., snapshots of the system variables are shown for different parameterization, differing only in the choice  $\beta$  and  $\sigma$ , configurations A, B, C and D are listed in Table 4.3., other parameters are shared between the configurations and are stated in Table 4.2.. Each configuration is expected to represent the current physiological state a single patient.

All following results are for a system with  $N = 200$  oscillators, and snapshots taken at time  $T_{\text{sim}} = 2000$ , the end of the integration time, and show the stationary values at that time point.

**Table 4.3.:** Specific  $\beta$ - $\sigma$  combinations to illustrate simulation results.

	A	B	C	D
$\beta$	$0.5\pi$	$0.58\pi$	$0.7\pi$	$0.5\pi$
$\sigma$	1.0	1.0	1.0	0.2

In Figure 4.7., the left-most columns depicts the coupling matrices for the organ layer  $\mathbf{K}^1$  followed by two columns showing the phase velocities for each oscillator  $\dot{\varphi}_i^{1,2}$  and two columns showing the oscillator phases each layer  $\varphi_i^{1,2}(T_{\text{sim}})$ . The right-most column shows the coupling matrix for the immune layer  $\mathbf{K}^2$ . Each layer is sorted first from lowest to highest frequency and secondary by lowest to highest phase for better clarity. Rows C and C' share the same parameterization but are different samples from the same initialization distributions.

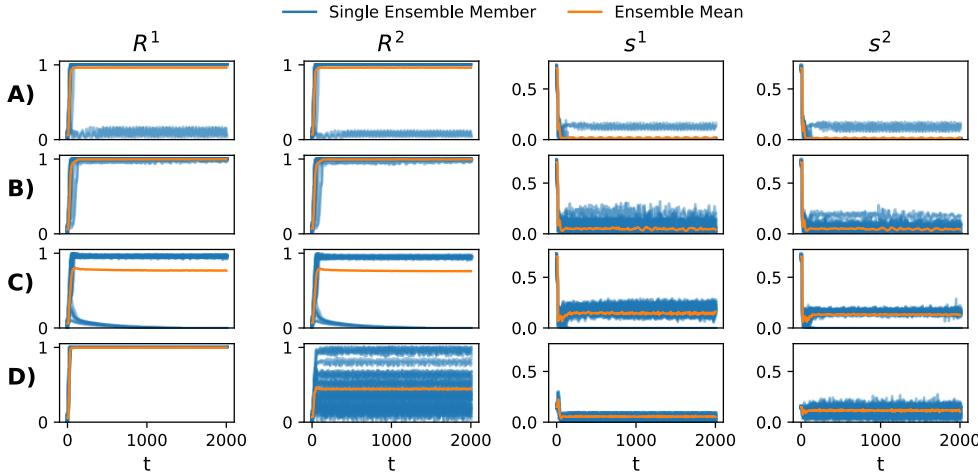


**Figure 4.7.:** Snapshots of different DNM parametrization and initialization. Configuration A can be regarded as healthy, with phases and frequencies being fully synchronized. In contrast, B and C are pathologic, due to their clustering in  $\dot{\phi}_i^1$ . Configuration C' corresponds to a vulnerable state, because of uniformly distributed phases (splay state). Lastly, D is regarded as resilient, since the phases exhibit clustering, but the frequencies  $\dot{\phi}_i^1$  are synchronized.

Row A in Figure 4.7. is fully synchronized/coherent since it not only has the frequencies synchronized but also the phases and can therefore interpreted as healthy. The coherence can also be seen in the fully homogeneous coupling matrices where both  $\mathbf{K}^{1,2}$  show the same coupling strength for every oscillator pair. The rows B and C in contrast show signs of a pathological state, here both the frequencies three and phases have four distinct clusters respectively. The clusters are also visible in the coupling matrices, where the coupling strength differs based on the cluster. The row for C', even though having the same parameterization as C, can be regarded vulnerable, since the phases uniformly distribute in the  $[0, 2\pi]$  interval ( $R^{1,2} = 0$ ), while the frequencies are synchronized. Coupling matrices for C' show traveling waves, which are characteristic for splay states. Observing different results for different initializations justifies the introduction of ensembles. Lastly row D shows a resilient state, where the phases are clustered while the frequencies are synchronized.

For the next result, the ensembles were introduced, every configuration of A, B, C, and D was simulated for  $M = 50$  different initializations over an interval of  $T_{\text{sim}} = 2000$ . The two left-most columns show the standard deviation of the mean phase velocities  $s^{1,2}$  for each ensemble member  $m$ . The plots show the temporal evolution of metrics for quantifying phase and frequency coherence, with the two right-most columns of Figure 4.8. show the temporal behavior of the Kuramoto Order Parameter for each individual ensemble member  $m \in 1, 2, \dots, M$ . Where lower values for  $R^{1,2}$  indicate decoherence in phases, with its minimum  $R^{1,2} = 0$  coincides with a splay state, and for  $s^{1,2}$  higher values indicate a larger amount of frequency decoherence and clustering.

which is stronger?

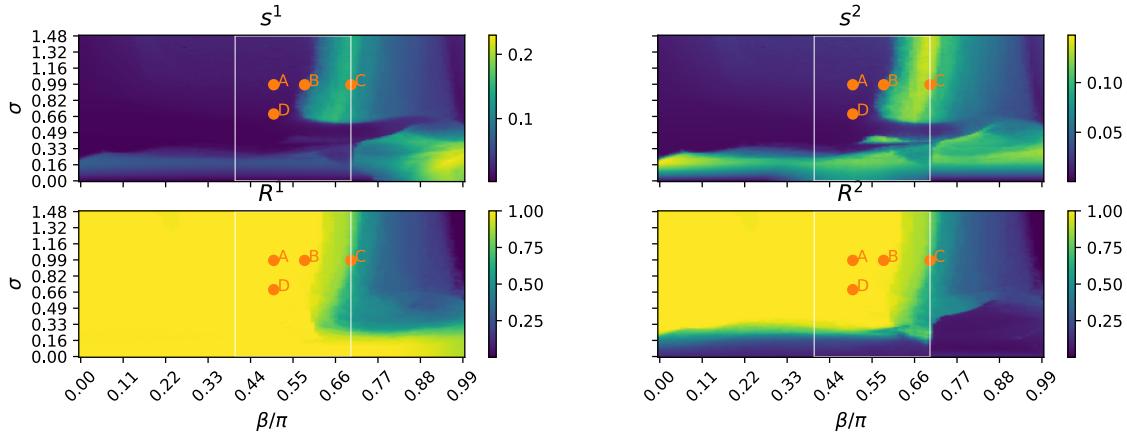


**Figure 4.8:** Transient and temporal evolution of the phase- and frequency-synchronization metrics  $R^{1,2}$  and  $s^{1,2}$ , for ensembles of the DNM for the configurations listed in Table 4.3.. Emphasizing the influence of  $\beta$  and  $\sigma$  on the systems synchronization behavior, and presenting different stable emergent system states.

Every ensemble in Figure 4.8. shows decoherence for early time-points, which is expected for randomly initialized variables, but changes relatively fast through a transient phase  $t \in [0.0, 200]$  into systematic stable behavior. The stable states align with the observations made for Figure 4.7., configuration A has, besides small jitter, mostly synchronized frequencies  $s^{1,2} \approx 0$ . Also the phases of configuration A are mostly synchronized with  $R^{1,2} \approx 1$ , only two initializations show decoherence and are oscillating between weak clustering and almost full incoherence. Medically this can be interpreted as a low risk of a dysregulated host response for an otherwise healthy response to the initial cytokine activation. For configuration B the amount of incoherence inside the ensemble is clearly visible, with  $s^{1,2}$  being positive and some more ensemble members exhibiting clustering, indicated by a Kuramoto Order Parameter slightly less than 1. In configuration C the incoherence is even more prominent, larger  $s^{1,2}$  and some ensemble members evolving into a splay state, i.e.  $R^{1,2} = 0$ . For configuration D the overall phase incoherence is again a bit less compared to C, and lower for the organ compared to the immune layer. The phases are coherent for the organ layer but seem almost chaotic for the immune layer, some are synchronized, while others are clustered, in a chimera or almost splay-like state. Over the whole simulation period, the coherency in the immune layer seems not to fully stabilize, rather oscillate around an attractor.

Mean erkären

Each of the configurations only differs in the parameter choices for  $\beta$  and  $\sigma$ , yet they evolve into unique and distinct system states. To put these four specific configurations into broader perspective, a grid of  $\beta$  and  $\sigma$  was simulated, in the interval  $\beta \in [0, 1]$  with a grid resolution of 0.01 and  $\sigma \in [0, 1.5]$  with a resolution of 0.015, creating a grid of 10,000 points. In Figure 4.9. the metric  $s^{1,2}$  is shown in the  $\beta - \sigma$  phase space for both layers in the first row, where brighter colors indicate a larger risk of frequency desynchronization, or risk of dysregulated immune response. The second row shows the ensemble mean over  $\bar{R}^{1,2}$ , i.e.  $\bar{R}^{1,2} = \frac{1}{M} \sum_m R_m^{1,2}$ , with  $M = 50$ , where darker colors indicate larger phase desynchronization. The white rectangle indicates the simulated region in [47],  $\beta \in [0.4, 0.7]$  and  $\sigma \in [0, 1.5]$  for reference. Coordinates of the configurations A, B, C, and D are also labeled.



**Figure 4.9.:** Phase space of the parameters  $\beta$  and  $\sigma$  and illustrating the broader picture their influence on the frequency and phase synchronization of the DNM. White rectangle indicates the grid-limits of the original publication [47].

Generally there is a similarity between phase and frequency desynchronization but no full equality, meaning there are parameter regions where the phase is synchronized and frequency desynchronized and vice versa. Another observation, that smaller values of  $\beta < 0.55$  correspond to less desynchronization and stronger coherence, which is in line with the medical interpretation of  $\beta$  where smaller values indicate a younger and more healthy biological age. When crossing a critical value of  $\beta_c \approx 0.55$  for the frequency and  $\beta_c \approx 0.6$  for the phases, the synchronization behavior suddenly changes and tends towards incoherence, clustering and pathological interpretations.

For small values of  $\sigma < 0.5$  the frequency synchronization and  $\sigma < 0.25$  for the phase synchronization, the behavior significantly differs between immune and organ layer. The immune layer tends to fully desynchronize, instead the organ layer only the frequency desynchronizes for larger  $\beta > 0.7$ . With larger values of  $\sigma > 0.5$  the dynamics more or less harmonize between layers and metrics and are mostly depend on  $\beta$ .

#### 4.4. Limitations and Research Direction

While the DNM provides a structured, functional and biologically inspired model sepsis for dynamics, several limitations must be acknowledged when interpreting its physiological relevance.

The fully connected intralayer topology (all oscillators inside a layer interact with each other), may not reflect actual cytokine networks behavior, where communication is spatially localized. Similarly the interlayer is modeled via uniform connections which treats infected areas not different than healthy areas, even though the communication patterns are most likely not the same for both cases.

In the original work, and adapted for this thesis, only a very small subspace has been simulated and interpreted [47]. For this selected subspace one can find regimes of synchronization,

phase clustering and frequency clustering, which allow physiological interpretation. But what about the unexplored space, does it not align with existing medical knowledge?

Central parameters like  $\beta$  the medical age,  $\sigma$  the interlayer coupling, the phase lags  $\alpha$ , and the timescale ratios  $\varepsilon$  do not correlate to any real observable quantity. This dimensionless formulation makes the model scale-free, which strengthens theoretical analysis but possibly weakens clinical translation.

The DNM is explicitly functional, where individual oscillators do not correspond to specific cells, instead the model captures aggregated behavior. This abstraction brings the benefit of computational tractability, compared to mechanistic cell simulation, and simplified system state interpretation, but it also brings drawbacks of limited ability to identify which biological processes drive the observed dynamics. It boils down to the question: Does the functional description capture the essential dynamics of sepsis, or does it produce patterns that resemble physiological behavior without true correspondence?

These limitations reveal a gap: while the DNM demonstrates rich theoretical behavior that can be interpreted through a physiological lens, the model has not yet been validated against real patient data. To bridge this gap between theoretical considerations and clinical utility, this work investigates whether the DNM can be grounded in observable patient trajectories. Specifically, can the abstract parameters ( $\beta, \sigma$ ) that govern the model behavior be inferred from real-world time-series data? And if so, does incorporating this physics-inspired structure provide advantages over purely data-driven approaches?

To summarize, the specific research questions include:

- 1) Usability of the DNM:** *How and to what extent can ML-determined trajectories of the DNM be used for detection and prediction, especially of critical infection states?*
- 2) Comparison with data-based approaches:** *How can the model-based predictions be compared with those of purely data-based approaches in terms of predictive power?*

The first question directly addresses the parameter interpretability problem. If  $\beta$  and  $\sigma$  cannot be reliably inferred from clinical observations, or if inferred values lack predictive power for outcomes, the models clinical relevance remains speculative. Conversely, if DNM parameters learned from data correlate with disease severity and trajectory, this would suggest physiological correspondence to the model parameters.

The second question tackles the functional vs. mechanistic limitation. Does embedding the DNM structure as an inductive bias improve predictions compared to black-box models that ignore physiological principles? Or does the model's abstraction level sacrifice too much biological detail to be useful? This comparison will reveal whether the DNM's theoretical framework translates to practical advantages in clinical prediction tasks.

## 4.5. Summary of the Dynamic Network Model

This chapter introduced the Dynamic Network Model as a functional, mesoscopic description of coordinated physiological activity during sepsis, modeling cellular cytokine-based communication. Based on adaptive Kuramoto-type oscillators arranged in a two-layer parenchymal-immune architecture, the model exhibits a range of emergent regimes like synchronization, clustering, chimera-like patterns, that may be interpreted physiological states. Key parameters such as the biological age  $\beta$  and interlayer coupling  $\sigma$  were shown to modulate these regimes. However, the gap between theoretical considerations and clinical validation remains. The following chapters address this gap directly by developing methods to infer DNM parameters from real patient data and evaluating whether this physics-based structure provides predictive advantages over purely empirical approaches.



# 5

## Method (Latent Dynamics Model)

This chapter introduces the methodological framework used to address the first research question as proposed in Section 4.4:

***Usability of the DNM: How and to what extent can the ML-determined trajectories of the DNM be used for detection and prediction, especially of critical infection states and mortality.***

As established in the previous chapter, the DNM provides a theoretical framework for understanding sepsis dynamics through coupled oscillator networks, but it has not yet been validated against real patient data. This chapter describes how the DNM can be embedded within a machine learning pipeline to enable parameter inference from clinical observations and probabilistic risk prediction.

Rather than predicting sepsis directly as a single binary outcome, proposed architecture decomposes the prediction task into two clinically meaningful quantities aligned with the Sepsis-3 definition. The Suspected Infection and increase in SOFA scores are predicted as proxies creating more nuanced and more interpretable prediction results.

The core idea is to embed the DNM into a learnable latent dynamical system. Here, a neural network learns to position patients into the  $(\beta, \sigma)$ -parameter phase space of the DNM and a recurrent neural network learns to predict the drift through that space based on observed clinical time series. The complete architecture, consisting of the DNM and additional auxiliary modules, will be referred to as the Latent Dynamics Model (LDM) from now on.

This chapter proceeds with Section 5.1, where the prediction task will be reiterated and the prediction strategy formalized. Desired prediction properties, together with the justification of modeling choices are also introduced here. Afterwards, in Section 5.2, the architecture will be discussed, focusing on what purpose each part serves and how it is integrated into the broader system.

### 5.1. Formalizing the Prediction Strategy

In automated clinical prediction systems, a patient is typically represented through their Electronic Health Record (EHR). The EHR aggregates multiple clinical variables, such as laboratory biomarker, for example from blood or urine tests, or physiological scores and,

further demographic information, like age and gender. Using the information that is available in the EHR until the prediction time-point  $t$ , the objective is to estimate the patients risk of developing sepsis at that time  $t$ . The following methodology will formalize the online-prediction, where newly arriving observations are continuously integrated into updated risk estimates. To use this prediction system in a clinical setting causality is important. Causality requires that for every prediction at time  $t$  only the information available up to that time-point can be used, and no future observations.

### 5.1.1. Patient Representation

Let  $t$  denote an observation time during a patients ICU-stay and the available EHR at that time consisting of  $D$  variables. After imputation of missing values, normalization, and encoding of non-numerical quantities, each variable  $\mu_j$  is mapped to a numerical value:

$$\mu_{t,j} \in \mathbb{R}, \quad j = 1, \dots, D \quad (5.1)$$

These values are collected into a column-vector:

$$\boldsymbol{\mu}_t = (\mu_{t,1}, \dots, \mu_{t,D})^T \in \mathbb{R}^D \quad (5.2)$$

where the superscript  $.^T$  denoting a transpose operation. The vector  $\boldsymbol{\mu}_t$  is fully describing the current physiological state of the ICU-patient at observation time point  $t$ . It is used as the feature vector, meaning it does not carry information that directly translate to the sepsis definition.

### 5.1.2. Modeling the Sepsis-3 Target

The goal is derive continuously updated estimates of sepsis risk based on newly arriving observations  $\boldsymbol{\mu}_t$  over time, with equally spaced and discrete time-steps  $t \in \{0, \dots, T\}$ . Following the Sepsis-3 definition, the onset of sepsis requires both suspected infection and acute multi-organ failure.

Defining the instantaneous *sepsis onset event*  $S_t \in \{0, 1\}$  as the occurrence of the Sepsis-3 criteria at time point  $t$  within the patients ICU stay as:

$$S_t := A_t \wedge I_t \quad (5.3)$$

Here  $A_t = \{\Delta O_t \geq 2\}$ , indicates an acute worsening in organ function, measured via a change in SOFA-score  $\Delta O_t = O_t - O_{\text{base}}$  with respect to some patient specific baseline SOFA-score  $O_{\text{base}}$ . The choice of  $O_{\text{base}}$  has to align with the Sepsis-3 definition, for example a 24 hours running minimum, the or  $O_{\text{base}} = O_{t-1}$ . The event  $I_t$  is an indicator for a Suspected Infection at time  $t$  defined according to the Sepsis-3 definition, spanning the 48 hours before and 24 hours after the documented SI-onset-time (see Section 2.2). Although the label  $I_t$  is defined retrospectively using a time window around the infection onset, this does not violate causality. The predictive model only conditions on  $\boldsymbol{\mu}_{0:t}$ , i.e., information available up to time

$t$ . Future observations are used exclusively for label construction during training and are not available at inference time.

Conditioned on the history of observations  $\mu_{0:t}$  the target probability is given by:

$$\Pr(S_t | \mu_{0:t}) = \Pr(A_t \cap I_t | \mu_{0:t}) \quad (5.4)$$

### 5.1.3. Heuristic Scoring and Risk Estimation

The direct estimation of the true conditional probability  $\Pr(S_t | \mu_{0:t})$  is computationally and statistically challenging due to the temporal dependency between the binary Sepsis-3 criteria. To make the prediction of the target probability more tractable but still connect the statistical estimation to the clinical definition several assumptions and modeling choices are introduced.

Importantly, all assumptions result in differentiable approximations of the real events or probabilities, enabling end-to-end learning of estimators through gradient-based methods.

The central assumption is that infection  $I_t$  and multi-organ failure  $A_t$  are conditionally independent:

$$\Pr(A_t \cap I_t | \mu_{0:t}) = \Pr(I_t | \mu_{0:t}) \Pr(A_t | \mu_{0:t}) \quad (5.5)$$

Clinically this assumption does not hold, since the majority multi-organ failures stem from an underlying infection, meaning they exhibit strong partial correlations. Yet this assumption is necessary because the DNM, which is an essential building block to the LDM, only captures organ failure risk irrespective of infection states and the independence allows treating both components separately for the prediction. Additionally, this separation improves interpretability, since each component can be analyzed individually.

As a second assumption, although the indication  $I_t$  is binary, the prediction target is a temporally smoothed version. The surrogate label  $\tilde{I}_t \in [0, 1]$  increases linearly in the 48 hours preceding the infection onset, it reaches maximum at onset, and it decays exponentially afterwards (24 hour window). This is mimicking temporal uncertainty of the diagnosis, for example due to delayed documentation and treatment effects such as antibiotic half-life.

Thus, the overall prediction requires two separate risk estimators:

$$\tilde{A}_t \approx \Pr(A_t | \mu_{0:t}) \text{ and } \tilde{I}_t \approx \bar{I}_t \quad (5.6)$$

Where  $\tilde{A}_t \in (0, 1)$  will be implemented as a heuristic risk scores, to conform differentiability, serving as approximation for the real event probabilities. Due to this implementation, the original prediction target has been converted from a calibrated probability to a *heuristic risk score*  $\tilde{S}_t$ :

$$\Pr(S_t | \mu_{0:t}) \approx \tilde{S}_t := \tilde{A}_t \tilde{I}_t \quad (5.7)$$

The interaction term  $\tilde{A}_t \tilde{I}_t$  mirrors their logical conjunction in the Sepsis-3 definition. It is important to note that  $\tilde{S}$  is **not a calibrated probability** but a heuristically derived and empirical risk score based on the Sepsis-3 definition, serving as a differentiable surrogate

for the Sepsis-3 sepsis onset criterion  $P(S_t | \mu_{0:t})$ . Larger values of  $\tilde{S}_t$  correspond to higher expected risk of sepsis outbreak.

#### 5.1.4. From EHR to Risk Scores

The high-dimensional EHR history  $\mu_{0:t}$  must now be condensed into these two clinically motivated statistics  $\tilde{A}_t$  and  $\tilde{I}_t$ . The LDM architecture implements two learned mappings:

**Infection risk estimation:** A data-driven module directly estimates infection risk from the EHR history:

$$\tilde{I}_t = f(\mu_{0:t}; \theta_f) \quad (5.8)$$

where  $f(\cdot; \theta_f)$  represents a neural network with learnable parameters  $\theta_f$  that will be specified in Section 5.2.

**Organ dysfunction estimation:** Rather than directly predicting  $\tilde{A}_t$  from the EHR, the LDM uses an intermediate representation, a latent SOFA-score estimate:

$$\hat{O}_t := g(\mu_{0:t}; \theta_g) \quad (5.9)$$

where  $\hat{O}_t$  denotes a latent, differentiable estimation for the true SOFA-score  $O_t$ . The function  $g(\cdot; \theta_g)$  represents a combined DNM pipeline, where  $\theta_g$  combines all learnable parameters of that pipeline. Again the function is further specified in Section 5.2.

Given two consecutive estimated SOFA-scores  $\hat{O}_{t-1}$  and  $\hat{O}_t$  a differentiable increase indicator  $\tilde{A}_t$  is calculated to indicate the event of organ failure:

$$\tilde{A}_t = o_{s,d}(\hat{O}_{t-1:t}) = \text{sigmoid}(s(\hat{O}_t - \hat{O}_{t-1} - d)) \quad (5.10)$$

The function  $o_{s,d}(\cdot)$  contains two globally learnable parameters,  $d$  a threshold and  $s$  a sharpness parameter. While the Sepsis-3 definition corresponds to a fixed threshold of  $d = 2$ , here  $d$  is treated as learnable to obtain a smooth, fully differentiable approximation of the discrete SOFA increase criterion and to account for uncertainty in baseline estimation. The choice of the function

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (5.11)$$

yields a monotonic indicator (larger SOFA increase  $\rightarrow$  more likely organ failure) while still being differentiable. The reason  $\tilde{A}_t$  is not a probability is caused by the sigmoid baseline of 0.5 when  $\hat{O}_t - \hat{O}_{t-1} = d$ , not when  $\Pr(A_t | \mu_{0:t}) = 0.5$ , making it an arbitrary monotonic transformation rather than a calibrated probability. This is fine for deep learning because gradient-based optimization only requires differentiability and monotonicity—the model learns to make  $\tilde{A}_t$  a useful risk indicator through end-to-end training on actual sepsis outcomes.

## 5.2. Architecture

The previous subsection explained how the sepsis onset target even  $S_t$  can be decomposed into the conjunction of suspected infection indication  $I_t$  and organ failure event  $A_t$  that itself can be calculated from two consecutive SOFA-scores  $O_{t-1:t}$ . The presented Latent Dynamics Model is designed to estimate the fundamental components  $\tilde{O}_t$  and  $\tilde{I}_t$  from a history of EHR  $\mu_{0:t}$  to derive the heuristic sepsis risk score  $\tilde{S}_t \approx S_t$  for individual patients. Each component is estimated by RNN module, e.g. Gated Recurrent Unit (GRU) or Long Short-Term Memory (LSTM), enabling continuous estimation updates based on newly arriving measurements.

The following subsection will introduce the individual modules which are fully differentiable functions with learnable parameters allowing for optimization via first order gradient descent methods. Starting with the estimator module for the suspected infection indication module in Section 5.2.1, followed the organ failure estimation module in Section 5.2.2 which includes the DNM to derive SOFA estimates and lastly an auxiliary regularization module in Section 5.2.3.

**Table 5.4.:** Summarization of notation used in the Latent Dynamics Model methodology.

SYMBOL	DESCRIPTION
$i, N$	Patient index and total patients
$t, T_i$	Time point and trajectory length
$\mu_t \in \mathbb{R}^D$	EHR vector with $D$ variables at time $t$
$\mu_{0:t}$	EHR history from time 0 to $t$
$S_t, A_t, I_t \in \{0, 1\}$	Binary sepsis onset, organ failure, and infection events
$O_t, \Delta O_t \in [0, \dots 24]$	SOFA score and change from baseline
$\bar{I}_t \in [0, 1]$	Continuous surrogate infection indicator
$\hat{O}_t \in \{0, \dots 24\}$	Estimated SOFA score
$o_{s,d}(\hat{O}_{t-1:t})$	Differentiable increase detection function
$\tilde{S}_t, \tilde{A}_t, \tilde{I}_t \in (0, 1)$	Predicted sepsis, organ failure, and infection risks
$z = (z_\beta, z_\sigma)$	Latent coordinates in DNM parameter space
$\hat{z}_t, \Delta \hat{z}_t$	Predicted latent position and change
$s^1(z)$	Synchronization measure (desynchronicity) in DNM
$h_t \in \mathbb{R}^h$	Hidden state vector
$f_\theta, g_\theta^e, g_\theta^r, d_\theta$	Infection indicator, encoder, recurrent, decoder modules
$\theta$	Learnable parameters
$\mathcal{L}_{\text{sepsis}}, \mathcal{L}_{\text{inf}}, \mathcal{L}_{\text{sofa}}$	Primary sepsis, infection, and SOFA losses
$\mathcal{L}_{\{\text{dec}, \text{spread}, \text{boundary}\}}$	Auxiliary decoder, latent-diversity, and -boundary losses
$\lambda_i$	Loss weight for component $i$
$B$	Mini-batch size

### 5.2.1. Infection Indicator Module

The first module of the LDM estimates the presence of a SI, represented by the continuous surrogate indicator  $\bar{I}_t$ , the module predicts a continuous surrogate infection risk  $\tilde{I}_t \in (0, 1)$ . Given  $N$  patient trajectories with  $T_i$  pairs of EHR vectors and ground truth SI-indicator each:

$$(\mu_{i,t}, \bar{I}_{i,t}), \quad i = 1 \dots N, \quad t = 1 \dots T_i \quad (5.12)$$

a parameterized nonlinear recurrent function

$$f_{\theta_f} : \mathbb{R}^D \times \mathbb{R}^{H_f} \rightarrow (0, 1) \times \mathbb{R}^{H_f} \quad (5.13)$$

is trained to map the patients physiological state represented by the EHR to an estimated risk of suspected infection:

$$(\tilde{I}_t, h_t^f) = f_{\theta_f}(\mu_t, h_{t-1}^f) \quad (5.14)$$

The hidden state  $h_t$  propagates temporal information through time. For the first time-step  $t = 0$  a learnable initial hidden state  $h_0^f$  is used.

The model is implemented as a RNN, illustrated in Figure 5.10.. At each timestep, a recurrent cell updates the hidden state, and a learned linear projection  $W_f h_t^f$ , with  $W_f \in \mathbb{R}^{1 \times H_f}$ , followed by sigmoid activation produces the infection risk estimate:

$$h_t = \text{RNN-Cell}_{\theta_f^{\text{rnn}}}(\mu_t, h_{t-1}^f) \quad (5.15)$$

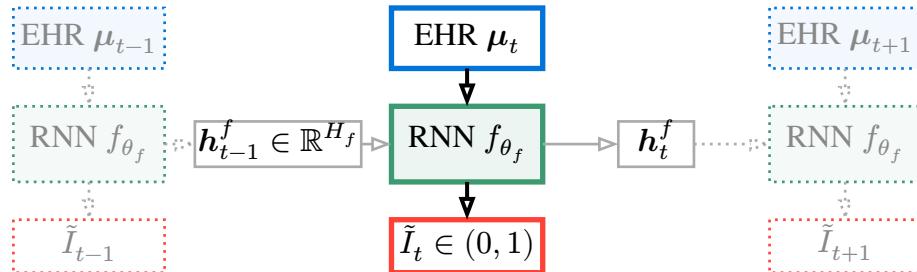
$$\tilde{I}_t = \text{sigmoid}(W_f h_t^f + b_f) \quad (5.16)$$

where  $\theta_f = \{\theta_f^{\text{rnn}}, W_f, b_f\}$  combines all learnable parameters, the bias  $b_f \in \mathbb{R}$  is a single scalar.

To fit the model, given a mini-batch if size  $B$ , Binary Cross Entropy (BCE) loss which measures the distance between true label  $\bar{I}_t$  and the predicted label  $\tilde{I}_t$ :

$$\mathcal{L}_{\text{inf}} = -\frac{1}{B} \sum_{i=1}^B \frac{1}{T_i} \sum_{t=1}^{T_i} \left[ \bar{I}_{i,t} \log(\tilde{I}_{i,t}) + (1 - \bar{I}_{i,t}) \log(1 - \tilde{I}_{i,t}) \right] \quad (5.17)$$

is minimized and thus the estimator provides a differentiable estimate of the surrogate infection activity.



**Figure 5.10.:** Schematic of the Infection Indicator Module architecture and rollout. The RNN process the EHR sequence  $\mu_{0:T}$  step-by-step, maintaining  $h_t$  to capture temporal dependencies, and outputs infection risk estimates  $\tilde{I}_t$  at each timestep.

## 5.2.2. SOFA Predictor Module

The complete SOFA predictor module  $g_\theta$  is composed two submodules, an initial-encoder  $g_\theta^e$  and a recurrent latent predictor  $g_\theta^r$ , each described below. The idea is to translate the physiological patient trajectory to a sequence of the DNM parameters  $\beta$  and  $\sigma$ , where the desynchronization of given parameter pairs should match the physiological organ failure. To begin with, Section 5.2.2.1 once more tries to strengthen the connection between organ failure and the DNM, followed by Section 5.2.2.2 presents how the EHR information is embedded evolved inside the DNM parameter space. Lastly Section 5.2.2.3 describes how computational cost can be significantly reduced by precomputing the DNM parameter space.

### 5.2.2.1. The DNM as SOFA Surrogate

Recalling that the pathological organ conditions within the DNM are characterized by frequency clustering in the parenchymal layer. The amount of frequency clustering is quantified by the ensemble average standard deviation of the mean phase velocity  $s^1$  (see Equation (4.15)). Since  $s^1$  monotonically increases with loss of frequency synchronization, it serves as an interpretable and natural surrogate for the SOFA-score. Increasing values of  $s^1$  indicate a higher SOFA-score and a worse condition of the patients organ system.

Numerical integration of the DNM equations for a given parameter pair  $z = (z_\beta, z_\sigma) = (\beta, \sigma)$  yields the corresponding SOFA approximation  $\hat{O}(z)$ . By taking the amount of desynchronicity at the end of the integration time  $s^1(T_{\text{sim}})$  at the coordinates of  $z$ , for readability the time argument is replaced by the parameter arguments  $z$ . Given a desynchronization measure  $s^z$ , the SOFA approximate is calculated using:

$$\hat{O}(z) = \left\lceil \frac{24 \cdot s^1(z)}{s^1_{\max}} \right\rceil = \left\lceil \frac{24 \cdot s^1(\beta, \sigma)}{s^1_{\max}} \right\rceil \quad (5.18)$$

Every other system parameter is assumed constant and chosen as listed in Table 4.2.. The space spanned by the two parameters is called the *latent space*, coordinate-pairs of that latent space are denoted  $z = (z_\beta, z_\sigma)$ . In this work only a predefined subspace of the entire  $(\beta, \sigma)$  plane is used. To normalize  $s^1$  to a  $[0, 1]$  range, and by this making it able to retrieve all 24 SOFA levels, the values of  $s^1$  are divided by the maximum value of the subspace  $s^1_{\max}$ . The rounding operation is used only for interpretability and evaluation; during training the normalized continuous  $s^1$  value is used.

The prediction strategy involves the mapping of individual EHR to the latent space, so that the ground truth SOFA aligns with the desynchronization measure of the latent coordinate. Based off of this initial location (and additional information), the patient will perform a trajectory through the latent space yielding step-by-step SOFA-score  $\hat{O}_{0:T}(z)$  estimates needed to calculate the heuristic organ failure statistic  $\tilde{A}_t$ .

### 5.2.2.2. Latent Parameter Dynamics

Focusing on a single patient, but omitting the  $i$  subscript for readability, with its first observation at time  $t = 1$ , an encoder connects the high-dimensional EHR to the dynamical regime of the DNM, a neural encoder:

$$g_{\theta_e}^e : \mathbb{R}^D \rightarrow \mathbb{R}^2 \times \mathbb{R}^{H_g} = \mathbb{R}^{2+H_g} \quad (5.19)$$

where the high dimensional patient state is mapped to a two-dimensional latent vector, and a  $H_g$ -dimensional hidden state.

$$(\hat{z}_0^{\text{raw}}, \mathbf{h}_0^g) = ((\hat{z}_{0,\beta}^{\text{raw}}, \hat{z}_{0,\sigma}^{\text{raw}}), \mathbf{h}_0^g) = g_{\theta_e}^e(\boldsymbol{\mu}_0) \quad (5.20)$$

This encoding locates the patient within a physiologically meaningful region of the DNM parameter space, which in context of the LDM is called the latent space. To keep latent coordinates in the predefined area they are ultimately transformed by:

$$\hat{z} = \text{sigmoid}(\hat{z}^{\text{raw}}) \cdot \left( \frac{\beta_{\max} - \beta_{\min}}{\sigma_{\max} - \sigma_{\min}} \right)^T + \left( \frac{\beta_{\min}}{\sigma_{\min}} \right)^T \quad (5.21)$$

Where  $\cdot$  is the element wise matrix multiplication. The latent coordinate  $\hat{z}_0$  provides the initial condition for short-term dynamical organ condition forecasting. As described in Section 5.2.2.1 the latent coordinates correspond to a DNM synchronization behavior and can therefore be directly interpreted as SOFA-score estimates ( $\hat{z} \rightarrow s^1(\hat{z}) \rightarrow \hat{O}(\hat{z})$ ).

In addition to the estimated parameter pair  $\hat{z}_0^{\text{raw}}$ , the encoder outputs another vector with dimension  $H_g \ll D$  that is a compressed representation of patient physiology relevant for short-term evolution of  $\hat{z}$ . This vector  $\mathbf{h}_0^g \in \mathbb{R}^{H_g}$  is the initial hidden space.

Since the heuristic SOFA risk  $\tilde{A}$  depends on the evolution of organ function  $\hat{O}_{0:t}^g$ , it is necessary to estimate not only the initial state  $\hat{z}_0$  but also its evolution. For this purpose a neural recurrent function:

$$g_{\theta}^r : \mathbb{R}^{D+2} \times \mathbb{R}^{H_g} \rightarrow \mathbb{R}^2 \times \mathbb{R}^{H_g} \quad (5.22)$$

is trained to propagate the latent DNM parameters forward in time.

This recurrent mechanism, conditioned on the hidden state  $\mathbf{h}_t^g$  and previous latent location  $\hat{z}_{t-1}^{\text{raw}}$ , captures how the underlying physiology influences the drift of the DNM parameters. From the previous hidden state and latent-position a recurrent cells updates the hidden state, followed by a linear down-projection  $\mathbf{W}_g \mathbf{h}_t^g$ , with  $\mathbf{W}_g \in \mathbb{R}^{2 \times H_g}$ , to receive the updated latent-position.

$$\mathbf{h}_t = \text{RNN-Cell}_{\theta_g^{\text{rnn}}}((\boldsymbol{\mu}_t, \hat{z}_{t-1}^{\text{raw}}), \mathbf{h}_{t-1}^g), \quad t = 2, \dots, T \quad (5.23)$$

$$\Delta \hat{z}_t^{\text{raw}} = (\mathbf{W}_g \mathbf{h}_t^g) \quad (5.24)$$

$$\hat{z}_t^{\text{raw}} = \hat{z}_{t-1}^{\text{raw}} + \Delta \hat{z}_t^{\text{raw}} \quad (5.25)$$

where  $\theta_g^r = \{\theta_g^{\text{rnn}}, \mathbf{W}\}_g \in \mathbb{R}\}$  combines all the learnable parameters. The down-projection does not have a bias-term so that no direction is inherently preferred.

Depending on the movement in the latent space the level of synchrony changes across the prediction horizon, which translates to the pathological evolution of patients. The online-prediction rollout is shown in figure Figure 5.11..

By predicting the movement in the latent space  $\Delta z_t$  instead of the raw parameters, smooth trajectories can be learned. For the latent sequence this is more desirable compared to the infection indicator, where jumps in predicted values do not matter as much.

To fit the functions, here the placement of latent points  $z$  is guided by a supervision signal through a Mean Squared Error (MSE) loss:

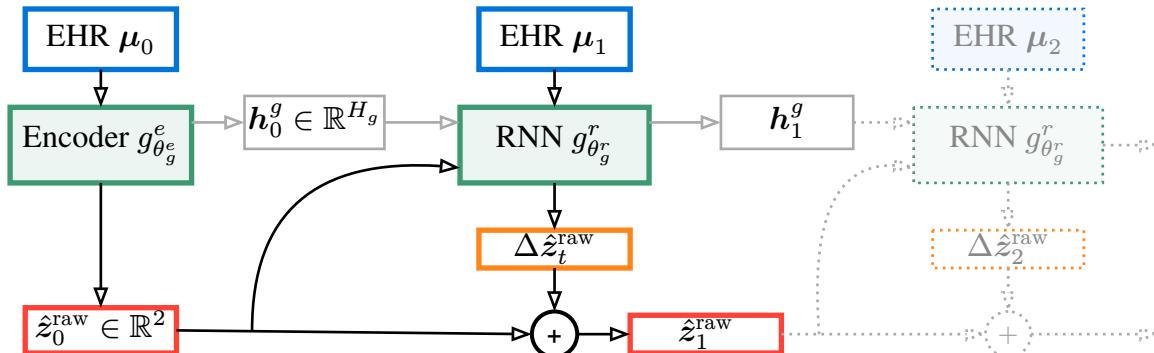
$$\mathcal{L}_{\text{sofa}} = \frac{1}{B} \sum_{i=1}^B \frac{1}{T_i} \sum_{t=1}^{T_i} w_{O_{i,t}} \cdot \left( \frac{O_{i,t}}{24} - \frac{s_{i,t}^1(\hat{z})}{s_{\max}^1} \right)^2 \quad (5.26)$$

where the class-balancing weight:

$$w_O = \log(1 + f_O^{-1}) \quad (5.27)$$

with  $f_O$  being the relative frequency of SOFA-scores  $O$ . This inverse-frequency weighting up-weights rare high SOFA-scores that are clinically critical but statistically underrepresented. Also notice that both parts, i.e. the continuous approximation (given by the desynchronicity) and ground truth are scaled to the interval  $[0, 1]$ .

Because gradients can flow backwards through the whole sequence, minimizing the loss can jointly train the encoder  $g_\theta^e$  and recurrent function  $g_\theta^r$ .



**Figure 5.11.:** Schematic of the online-prediction rollout by the SOFA Predictor Module. The Encoder  $g_{\theta_g}^e$ , generates the initial latent position  $\hat{z}_0^{\text{raw}}$  based on the first observed EHR. Afterwards, the RNN processes the following EHR sequence  $\mu_{0:T}$  step-by-step, maintaining  $h_t$  to capture temporal dependencies, and outputs the change in latent position  $\Delta\hat{z}_t^{\text{raw}}$  at each timestep. The new position is the sum of the previous position and its update  $\hat{z}_{t-1}^{\text{raw}} + \Delta\hat{z}_t^{\text{raw}}$

Raws

### 5.2.2.3. Latent Lookup

Intuitively one would numerically integrate the DNM every estimate  $\hat{z}$  to receive the  $s^1(\hat{z})$ -metric for the continuous space in  $(\beta, \sigma)$ . This approach is taken in Neural Differential

Equations [59] and Physics Informed Neural Networks [60] where gradients are typically backpropagated through the ODE integration to their input parameters ( $\beta, \sigma$  in this case). Practically, in case of the DNM this is hardly tractable, since the integration is computationally intensive and gradients are prone to vanish over the large integration time and ensemble setup of the DNM.

To address these challenges, the LDM uses a fully differentiable precomputing and caching methodology that still provides meaningful gradients and simultaneously reduces the computational burden. For that, the continuous latent space has been quantized to a discrete and regular grid, with the metric precomputed for each coordinate pair in the predefined subspace. The space is limited to the intervals  $\beta \in [0.4\pi, 0.7\pi]$  and  $\sigma \in [0.0, 1.5]$  (the phase space of the original publication [47]). To retrieve values that do not lie exactly on a grid-points, localized soft interpolation is used to derive differentiable synchronicity approximation values.

For an estimated coordinate pair  $\hat{z} = (\hat{z}_\beta, \hat{z}_\sigma)$  in the continuous  $(\beta, \sigma)$ -space the quantized metrics are interpolated by smoothing nearby quantization points with a Gaussian-like kernel, which is illustrated in Figure 5.12..

To enable gradient-based optimization, i.e. being differentiable, the lookup of nearby points  $z'$  combines two mechanisms. Firstly, a straight-through estimator [61] for the discrete voxel indexing operation, allowing gradients to flow as if the rounding were identity.

$$\tilde{z} = \hat{z} + \text{stop\_grad}(\lfloor \hat{z} \rfloor - \hat{z}) \quad (5.28)$$

In the forward pass this equals the rounded value for lookup, in the backwards pass the stop\_grad operation blocks gradients from the rounding, so the gradient flows as if no rounding occurred. Secondly, a differentiable softmax interpolation over neighboring grid points. The nearby points are selected by a rectangular kernel around the closest quantized point  $\tilde{z}$ . Given a kernel-size  $k$  the approximated values is calculated by:

$$\tilde{s}^1(\tilde{z}) = \sum_{z' \in \mathcal{N}_{k \times k}(\tilde{z})} \text{softmax}\left(-\frac{\|\tilde{z} - z'\|^2}{T_d}\right) s^1(z') \quad (5.29)$$

with softmax for  $K = k \cdot k$  neighboring points, where  $k$  is an odd number  $> 1$ . Here,  $T_d \in \mathbb{R}_{>0}$  is a learnable temperature parameter which controls the sharpness of the smoothing, with larger values producing stronger smoothing and smaller values converging to the value of the closest point  $\tilde{z}$  exclusively. This allows the model to adjust the interpolation sharpness during training, potentially using broad smoothing early on for exploration and sharpening later for precision.

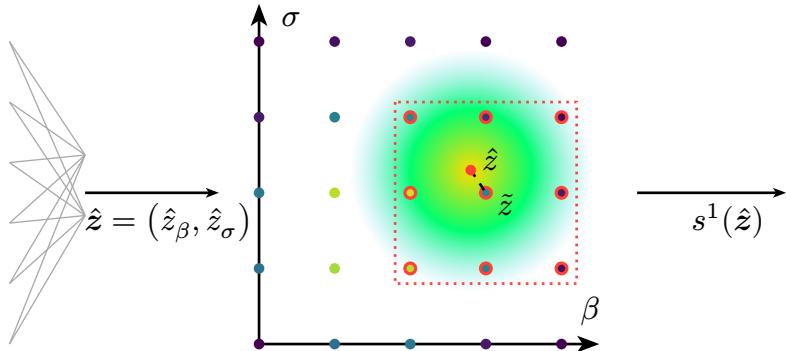
While the squared distances ( $\|\tilde{z} - z'\|^2$ ) receive exponentially more weight, the softmax operation normalizes the weights to 1, creating a proper convex combination of weights.

$$\text{softmax}(\mathbf{x})_j = \frac{e^{x_j}}{\sum_{k=1}^K e^{x_k}}, \quad \text{for } j = 1, \dots, K \quad (5.30)$$

The  $K$  neighboring points can be calculated via:

$$\mathcal{N}_{k \times k}(\tilde{z}) = \left\{ (\tilde{z}_\beta + i \cdot \beta_{\text{step size}}, \tilde{z}_\sigma + j \cdot \sigma_{\text{step size}}) \mid \right. \quad (5.31)$$

$$\left. i, j \in -\left(\frac{k-1}{2}\right), \dots, -1, 0, 1, \dots, \left(\frac{k-1}{2}\right) \right\} \quad (5.31)$$



**Figure 5.12.:** Quantized latent lookup of precomputed synchronization metrics. Point colors represent the amount of desynchronization  $s^1$  in the parenchymal layer. Neighboring points, the  $z' \in \mathcal{N}_{3 \times 3}(\tilde{z})$  sub-grid, indicated by the red outlines and the red rectangle around  $\tilde{z}$ , are used smoothed using a Gaussian-like kernel, represented by the color gradient around estimation point  $\tilde{z}$ . This allows continuous interpolation the parameter space.

This quantization strategy, called *latent lookup*<sup>2</sup> is closely related to Finite Scalar Quantization (FSQ) [62], used in Dreamer-V3 [63] for example. In contrast to this presented latent lookup, the latent coordinates in Dreamer-V3 do not have prior semantic meaning associated with them. Both allow for differentiable quantization, with details on the latent lookup implementation, including grid-resolution and kernel size, can be found in Section 6.2.

### 5.2.3. Decoder

As shown in the visualization of the DNM phase space in Figure 4.9. multiple latent coordinates  $z$  result in the same amount of desynchronization, since different physiological states share the same SOFA level. But when different physiological states have a common SOFA-score, their latent representations should be different and unique to that exact physiological state. This should enable to distinguish different triggers of the organ failure inside the latent space, similarly to how it is possible to distinguish the different triggers from the EHR.

In a classical Auto-Encoder [64] setting, to encourage a semantically structured latent space, a decoder module is added as an auxiliary regularization component. A neural decoder network:

$$d_{\theta_d} : \mathbb{R}^2 \rightarrow \mathbb{R}^D \quad (5.32)$$

attempts to reconstruct the original EHR features from the latent representation, the resulting desynchronicity of that latent coordinate and the heuristic risk measures:

<sup>2</sup>Implementation is available at [https://github.com/unartig/sepsis\\_osc/blob/main/src/sepsis\\_osc/lsm/lookup.py](https://github.com/unartig/sepsis_osc/blob/main/src/sepsis_osc/lsm/lookup.py)

$$\hat{\mu}_t = d_{\theta_d}(\hat{z}_t) \quad (5.33)$$

This way the decoder only learns to disentangle the latent coordinates in  $\hat{z}_t$  based on ground future EHRs  $\mu_t$ . The module is trained using a supervised loss:

$$\mathcal{L}_{\text{dec}} = \frac{1}{B} \sum_{i=1}^B \frac{1}{T_i} \sum_{t=0}^{T_i-1} (\mu_{i,t} - \hat{\mu}_{i,t})^2 \quad (5.34)$$

This serves as regularization because the reconstruction objective forces the latent space to maintain a structured organization where physiologically distinct states are positioned into different regions, rather than allowing arbitrary latent encodings.

This latent regularization is motivated by *Representation Learning* [64] and ensures that nearby points in the latent  $(\beta, \sigma)$ -space correspond to physiologically similar patient states. It should help the encoder  $g_{\theta_g}^e$  to learn a meaningful alignment between EHR-derived latent embeddings and the dynamical DNM landscape. Using this regularization the latent encoder  $g_{\theta_g}^e$  and the recurrent predictor  $g_{\theta_g}^r$  are encouraged to map temporally consecutive to spatially near latent coordinates, since it is expected that consecutive EHRs do not exhibit drastic changes. Leading smooth patient trajectories through the latent space.



**Figure 5.13.:** Schematic of the data flow in the decoder module. The decoder network  $d_{\theta_d}$  tries to reconstruct every latent coordinate pair  $z_t$  to the original EHR features  $\mu_t$ . This auxiliary component encourages semantically structured latent representations: physiologically similar patient states map to nearby points in the  $(\beta, \sigma)$ -space, while different triggers of organ failure occupy distinct regions.

#### 5.2.4. Combining Infection and Acute Change Signals

The complete LDM, shown in Figure 5.14., is trained jointly by combining the previously introduced Infection Indicator Module  $f_{\theta_f}$  and the SOFA prediction module  $g_{\theta_g}$ . The output of these modules yield the components  $\tilde{O}_t$ , from which  $\tilde{A}_t$  can be derived (Equation (5.10)) and  $\tilde{I}_t$ .

Because positive labels may be temporally windowed around the true onset of sepsis  $S_t$ , the estimated sepsis risk score is computed via causal smoothing:

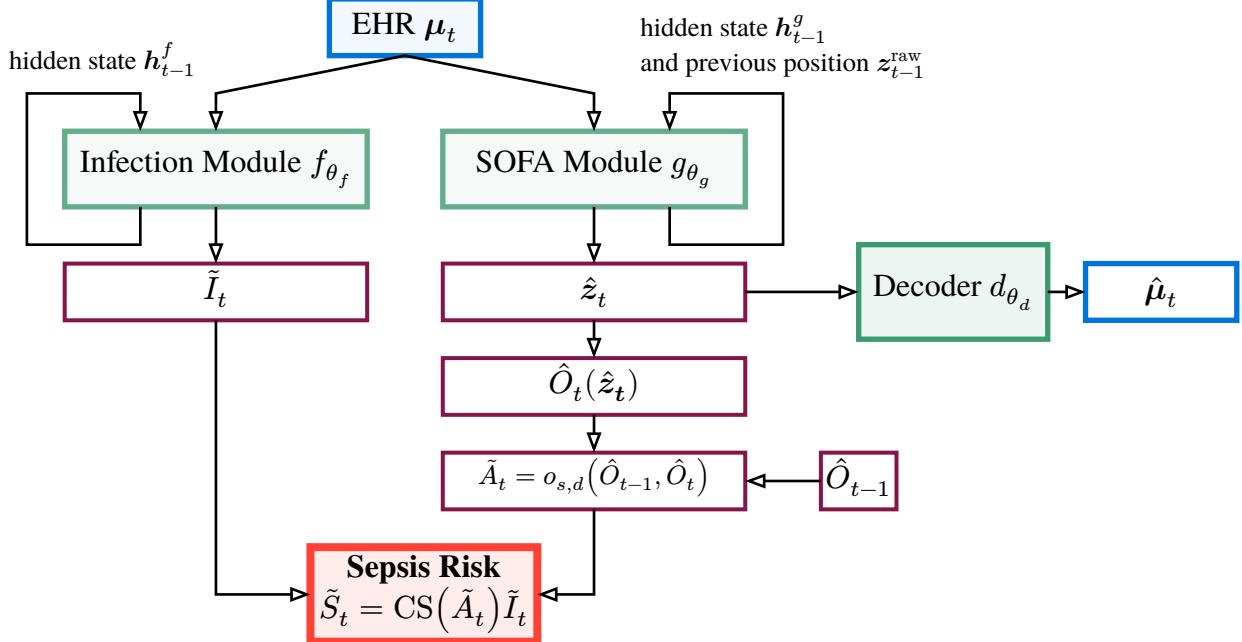
$$\tilde{S}_t = \text{CS}(\tilde{A}_t) \cdot \tilde{I}_t \quad (5.35)$$

where  $\text{CS}(\cdot)$  denotes a causal smoothing operator that maintains elevated predictions in the time-steps preceding sepsis onset. The causal smoothing operation is defined as:

$$\text{CS}(x_t) = \sum_{\tau=0}^r w_\tau \cdot x_{t-\tau}, \quad w_\tau = \frac{e^{-\alpha\tau}}{\sum_{k=0}^r e^{-\alpha k}} \quad (5.36)$$

with radius  $r$  is a hyper-parameter controlling the temporal radius of the smoothing window, and  $\alpha$  a learnable decay parameter controlling the length and shape of the smoothing kernel. To handle the sequence boundaries  $x_{t-\tau} = 0$  for  $t - \tau < 0$ .

This smoothing ensures that organ failure predictions remain elevated during the causal window preceding sepsis onset, matching the clinical reality that organ dysfunction typically precedes documented sepsis.



**Figure 5.14.:** Complete LDM architecture with three main components. The Infection Module  $f_{\theta_f}$  and SOFA Module  $g_{\theta_g}$  process EHR data  $\mu_t$  through recurrent networks to estimate infection level  $\tilde{I}_t$  and latent coordinates  $\hat{z}_t$  respectively. The latent coordinates map to organ failure  $\hat{O}_t$ , from which acute changes  $\tilde{A}_t$  are computed using consecutive predictions. The heuristic organ failure risk is assumed to be 0 for the initial time step. The Decoder  $d_{\theta_d}$  reconstructs EHR features  $\mu_t$  from latent coordinates, regularizing the latent space to maintain clinically meaningful structure. Final sepsis risk  $S_t$  combines infection and acute change signals.

### 5.3. Training Objective and Auxiliary Losses

Besides the losses already presented, to guide the training process multiple auxiliary losses are used and introduced in the following.

#### Primary Sepsis Prediction Loss

The main training signal aligns the heuristic sepsis score  $\tilde{S}_t$  with ground truth sepsis labels:

$$\mathcal{L}_{\text{sepsis}} = -\frac{1}{B} \sum_{i=1}^B \frac{1}{T_i} \sum_{t=1}^{T_i} [S_{i,t} \log(\tilde{S}_{i,t}) + (1 - S_{i,t}) \log(1 - \tilde{S}_{i,t})] \quad (5.37)$$

using the BCE.

### Latent Space Regularization

To prevent collapse and ensure diverse latent representations the following loss is introduced:

$$\mathcal{L}_{\text{spread}} = -\log(\det(\text{Cov}(\hat{\mathbf{Z}}))) \quad (5.38)$$

where  $\hat{\mathbf{Z}} \in \mathbb{R}^{2 \times B \cdot T}$  collects all predicted latent coordinates of a batch.  $\text{Cov}(\cdot)$  computes the sample covariance matrix.

The loss is minimized when the *generalized variance* [65] of the latent dimensions  $\beta$  and  $\sigma$ . The generalized variance roughly measures the density of distributions and increases they become more dense, the loss  $\mathcal{L}_{\text{spread}}$  therefore encourages a larger spread inside the latent space.

### Latent Space Regularization

In order to keep the predicted latent inside the predefined area, they will be discouraged to move too close to the edges:

$$\mathcal{L}_{\text{boundary}} = \text{ReLU}(f - \text{sigmoid}(z_t^{\text{raw}})) + \text{ReLU}(\text{sigmoid}(z_t^{\text{raw}} - (1 - f))) \quad (5.39)$$

with  $f \in (0, 0.5)$  sets a boundary threshold as a fraction of the space, creating a “penalty buffer” that discourages latent variables from entering the outer  $f$ -percent of the space near the edges..

#### 5.3.1. Combined Objective

The complete LDM<sup>3</sup> is trained jointly by optimizing all components with the weighted total loss:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{inf}} \mathcal{L}_{\text{inf}} + \lambda_{\text{sofa}} \mathcal{L}_{\text{sofa}} + \lambda_{\text{dec}} \mathcal{L}_{\text{dec}} + \lambda_{\text{sepsis}} \mathcal{L}_{\text{sepsis}} + \quad (5.40)$$

$$\lambda_{\text{spread}} \mathcal{L}_{\text{spread}} + \lambda_{\text{boundary}} \mathcal{L}_{\text{boundary}} \quad (5.40)$$

The loss weights  $\lambda$  balance the contribution of each objective during training. The primary sepsis prediction loss  $\mathcal{L}_{\text{sepsis}}$  provides the main learning objective aligned with the clinical task, while component losses  $\mathcal{L}_{\text{inf}}$  and  $\mathcal{L}_{\text{sofa}}$  ensure accurate estimation of the underlying infection and organ failure indicators. The auxiliary losses ( $\mathcal{L}_{\text{dec}}, \mathcal{L}_{\text{spread}}$ ) regularize the latent space structure to improve generalization and interpretability. Specific values for the loss weights  $\lambda$  and other hyperparameters are reported in Section 6.2.

Table 5.5. provides an overview of all loss components, their purpose, and the modules they supervise.

**Table 5.5.:** Overview of loss components in the LDM training objective.

Loss	Type	Purpose	Supervises
$\mathcal{L}_{\text{sepsis}}$	BCE	Primary sepsis prediction	$f_{\theta_f}, g_{\theta_g}^e, g_{\theta_g}^r$
$\mathcal{L}_{\text{sofa}}$	Weighted MSE	SOFA estimation	$g_{\theta_g}^e, g_{\theta_g}^r$

<sup>3</sup>Implementation of the LDM components is available at [https://github.com/unartig/sepsis\\_osc/tree/main/src/sepsis\\_osc/lmd](https://github.com/unartig/sepsis_osc/tree/main/src/sepsis_osc/lmd)

$\mathcal{L}_{\text{inf}}$	BCE	Infection indicator	$f_{\theta_f}$
$\mathcal{L}_{\text{spread}}$	Covariance	Latent diversity	$g_{\theta_g^e}, g_{\theta_g^r}$
$\mathcal{L}_{\text{boundary}}$	Positional	Latent Space	$g_{\theta_g^e}, g_{\theta_g^r}$
$\mathcal{L}_{\text{dec}}$	MSE	Latent semantics	$d_{\theta_d}, (g_{\theta_g^e}, g_{\theta_g^r})$

## 5.4. Inference

At inference time, the LDM operates as a continuous monitoring system for ICU patients, providing real-time risk assessment from admission through the entire ICU stay. Upon patient admission to the ICU, and once initial laboratory measurements are available, the first EHR observation  $\mu_0$  is processed by both the Infection Indicator module and the latent encoder. The infection indicator  $f_{\theta_f}$  produces an initial infection risk estimate  $\tilde{I}_0$  and hidden state  $h_0^f$ . The latent encoder  $g_{\theta_g^e}$  maps the  $\mu_0$  to the initial latent coordinates  $\hat{z}_0$ . Deriving the synchronicity measure  $s_0^1(\hat{z})$  from the coordinates provides an immediate indication of organ system functionality.

This initialization establishes the patients baseline physiological state within the DNM parameter space and provides initial risk indicators. The hidden states  $h_0^f$  and  $h_0^g$  are saved to enable temporal continuity in subsequent predictions. Triggered by newly arriving measurements or at regular hourly intervals, the system performs sequential updates. Updated EHRs  $\mu_t$  are processed by the recurrent modules  $f_{\theta_f}$  and  $g_{\theta_g^r}$  generating updated estimates on the infection risk and organ system state  $\tilde{I}_t$  and  $\hat{z}_t$ . From the history of the latent trajectory  $\hat{z}_{0:t}$  the acute risk of organ failure  $\tilde{A}_t$  is calculated and the risk of sepsis estimated  $\tilde{S}_t$ . This process is run until the patient leaves the ICU.

Overall, at inference time, the LDM provides multiple clinically interpretable indicators at each timestep:

- $\tilde{I}_t \in (0, 1)$ : Current infection likelihood
- $s_t^1(\hat{z}) \in [0, 1]$ : Organ system desynchronization (proxy for SOFA score)
- $\tilde{A}_t \in (0, 1)$ : Acute organ failure risk (recent worsening)
- $\tilde{S}_t \in (0, 1)$ : Overall sepsis risk (primary alert signal)

These outputs allow clinicians to not only assess overall sepsis risk but also understand the contributing factors, whether the risk stems primarily from suspected infection, acute organ deterioration, or both. Additionally, the latent trajectory  $s_{0:t}^1(\hat{z})$  through the DNM parameter space provides interpretable visualization of the patients physiological evolution over time.

## 5.5. Assessing the Prediction Performance

In order to qualitatively assess the prediction performance of sepsis prediction models two theoretically grounded metrics will be introduced. The prediction of a patient developing sepsis vs. no sepsis is a binary decision problem based off of the continuous estimated

heuristic sepsis risk  $\tilde{S}_t$ . Given an estimated risk  $\tilde{S}_t$  and a decision threshold  $\tau \in [0, 1]$ , the decision if a prediction value counts as septic is given by the rule:

$$\delta(\tilde{S}_t) = \mathbb{I}(\tilde{S}_t > \tau) \quad (5.41)$$

where  $\mathbb{I}(\cdot)$  is 1 when the condition is met and 0 otherwise. For different choices of  $\tau$  the decision rule can be applied and yield different ratios of:

- **True Positives (TP)** where truth and estimation are 1
- **False Positives (FP)** where truth is 0 and estimation 1
- **True Negatives (TN)** where truth and estimation are 0
- **False Negatives (FN)** where truth is 1 and estimation 0,

from these one can calculate the True Positive Rate (TPR) (also called sensitivity):

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5.42)$$

and the False Positive Rate (FPR):

$$\text{FPR} = \frac{\text{FP}}{\text{TP} + \text{FN}} \quad (5.43)$$

Sweeping the decision boundary  $\tau$  from 0 to 1 and plotting the corresponding implicit function TPR vs FPR creates the *receiver operating characteristic* or *ROC* curve. A prediction system operating at chance will have exhibit a diagonal line from (0 FPR, 0 TPR) to (1 FPR, 1 TPR). Everything above that diagonal indicates better predictions than chance, with an optimal predictor “hugging” the left axis until the point (0 FPR, 1 TPR) followed by hugging the top axis. The quality of the whole curve can be summarized to a single number, the area under the curve, called AUROC, where larger values  $\leq 1$  indicate better prediction performance.

When trying to predict rare events, meaning sparse positive against lots of negative events the FPR can become small and thus little informative. In these cases one commonly plots the *precision*:

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5.44)$$

against the *recall*:

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5.45)$$

creating the *precision recall curve* or PRC, where an optimal predictor hugs the top right. Also this curve can be summarized by its area to the AUPRC metric, where larger values indicate better performance [66].

Traditionally, the AUPRC is referred to as the more appropriate metric for imbalanced prediction tasks. Though, recent research suggests the AUROC as a more reliable metric for use cases with elevated FN costs, such as the increased mortality risk in false or delayed sepsis diagnoses [67]. Together the AUROC and AUPRC are the commonly reported perfor-

mance metrics used in the sepsis prediction literature [31] and will be used to compare the performance between the LDM and a baseline approach.

## 5.6. Summary of Methods

This chapter introduced the proposed model for short-term and interpretable risk prediction of developing sepsis for ICU patients, referred to as Latent Dynamics Model. Starting from the formal task definition, the full processing pipeline and detailed the architecture of the encoder, recurrent latent dynamics module, decoder, and the infection-indicator classifier have been presented. A key component of the approach is the integration of the functional DNM into the latent dynamics, enabling physiologically meaningful and interpretable temporal modeling.

The training losses, including auxiliary losses, used for each component were defined and explained how they contribute to the overall optimization objective. Additionally, chapter explained how the LDM can be used to support clinical monitoring of patients. Lastly, the two metrics AUROC and AUPRC were introduced to assess the prediction performance.

The next Section 6 presents an experiment where the LDM is trained using a widely used data-source ICU in order to benchmark sepsis prediction capabilities. Therefore, it presents the exact task and cohort definitions and LDM parameterization as well as the training procedure. The relevant evaluation metrics AUROC and AUPRC are used to assess the predictive performance and compare to existing baseline methods.



# 6

## Experiment

To evaluate whether embedding the Dynamic Network Model improves short-term sepsis prediction, the Latent Dynamics Model (see Section 5) was trained and evaluated using real-world medical data. This chapter presents the complete experimental setup and results, including data basis (data source, cohort selection, preprocessing in Section 6.1), the prediction task, and implementation and training details in Section 6.2 and Section 6.2.1. Finally, the chapter will end with the discussion of experimental results using the previously established metrics AUROC and AUPRC, along with qualitative analyses of individual patient trajectories through the DNM parameter space.

### 6.1. Data

This study relies exclusively on the Medical Information Mart for Intensive Care-IV database (version 2.3) [37]. MIMIC database series contains EHR information capturing day-to-day clinical routines, including patient measurements, orders, diagnoses, procedures, treatments, and free-text clinical notes. All included EHRs were recorded at Beth Israel Deaconess Medical Center in Boston, America between 2008 and 2022. Every part of the data has been de-identified and is publicly available to support research in electronic healthcare applications, with special focus on intensive care. While applications trained on MIMIC databases are known to have limited generalization to other data-sources and real-world settings, they remain the default open-data resource for developing sepsis prediction systems [30], [68].

#### 6.1.1. Cohort Definition, Feature Choice and Preprocessing

To derive a cohort from raw data and preprocess clinical features, the YAIB framework is employed [38]. YAIB standardizes cohort definition, feature derivation and data preprocessing for retrospective ICU studies across different publicly available databases. It additionally provides benchmark results for common ICU prediction task, including the online prediction of sepsis. For this work, every step from the sepsis and cohort definition, feature choices to the data preprocessing, is adopted from their methodology [38] to enable direct comparison of prediction results.

Their definition closely follows the Sepsis-3 criteria [4]:

“ The onset of sepsis was defined using the Sepsis-3 criteria (Singer et al., 2016), which defines sepsis as organ dysfunction due to infection. Following guidance from the

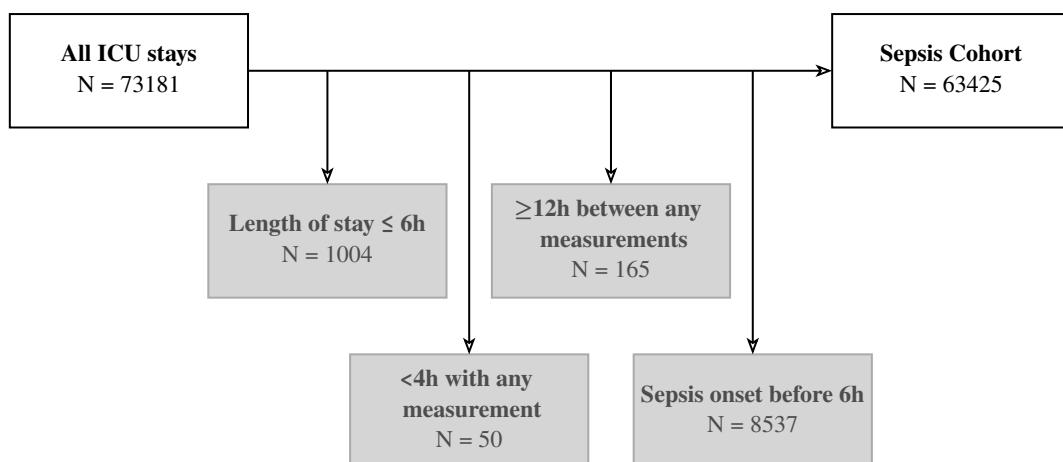
original authors of Sepsis-3 (Seymour et al., 2016), organ dysfunction was defined as an increase in SOFA score  $\geq 2$  points compared to the lowest value over the last 24 hours. Suspicion of infection was defined as the simultaneous use of antibiotics and culture of body fluids. The time of sepsis onset was defined as the first time of organ dysfunction within 48 hours before and 24 hours after suspicion of infection. Time of suspicion was defined as the earlier antibiotic initiation or culture request. Antibiotics and culture were considered concomitant if the culture was requested  $\leq 24$  hours after antibiotic initiation or if antibiotics were started  $\leq 72$  hours after the culture was sent to the lab. Where available, antibiotic treatment was inferred from administration records; otherwise, we used prescription data. To exclude prophylactic antibiotics, we required that antibiotics were administered continuously for  $\geq 3$  days (Reyna et al., 2019). Antibiotic treatment was considered continuous if an antibiotic was administered once every 24 hours for 3 days (or until death) or was prescribed for the entire time spent in the ICU “

### 6.1.2. Cohort Selection

The cohort includes all adult patients (age at admission  $\geq 18$ ,  $N = 73,181$ ). To ensure data volume and quality, patients meeting any of the following criteria were excluded:

- Less than six hours spent in the ICU.
- Less than four separate hours across the entire stay where at least one feature was measured.
- Any time interval of  $\geq 12$  consecutive hours throughout the stay during which no feature was measured.
- Sepsis onset before the 6th hour in the ICU.

Applying these criteria resulted in a final cohort size of  $N = 63,425$  patients. The selection process with corresponding exclusion numbers is shown in Figure 6.15..



**Figure 6.15.:** Cohort selection and exclusion process.

### 6.1.3. Cohort Characteristics

Table 6.6. presents the demographic and clinical characteristics of the final cohort, stratified by sepsis status according to the Sepsis-3 criteria. Of the 63,425 patients included, 3,320 (5.2%) met the criteria for sepsis. Sepsis-positive patients exhibited notably higher disease severity, with a median maximum SOFA score of 5.0 compared to 4.0 in sepsis-negative patients, and substantially higher hospital mortality (26.5% vs 6.6%). Additionally, septic patients had significantly longer Length Of Stay (LOS) than non-septic patients (median 335.1 hours vs 150.3 hours). Median time to sepsis onset was 13 hours (Inter Quantile Range (IQR) (25%-75%): 8–34).

Both groups were similar in terms of demographic characteristics, including age (median 65 years), sex distribution (approximately 55% male), and weight at admission (median 77.6 kg). Most patients in both groups were white (63.6% overall) and had medical admissions (71.0% overall), though sepsis-positive patients had a higher proportion of medical admissions (84.8% vs 70.2%).

**Table 6.6.:** Characteristics and demographics of the cohort. Numerical variables are summarized by *median [IQR 25 - 75]* and numerical variables by incidence (%)

CHARACTERISTIC	ALL PATIENTS	SEP-3 POSITIVE	SEP-3 NEGATIVE
<b>Demographics</b>			
N	63425 (100.0)	3320 (5.2)	60105 (94.8)
Male	35170 (55.5)	1881 (56.7)	33289 (55.4)
Age at admission	65.0 (53.0–76.0)	65.0 (54.0–76.0)	65.0 (53.0–76.0)
Weight at admission	77.6 (65.1–92.3)	77.6 (65.6–94.0)	77.6 (65.0–92.2)
<b>Clinical Outcomes</b>			
SOFA median	3.0 (1.0–5.0)	3.0 (1.0–5.0)	3.0 (1.0–5.0)
SOFA max	4.0 (2.0–6.0)	5.0 (4.0–8.0)	4.0 (2.0–6.0)
Hospital LOS hours	157.7 (92.8–268.9)	335.1 (194.2–548.6)	150.3 (90.9–256.0)
Hospital Mortality	4828 (7.6)	879 (26.5)	3949 (6.6)
Sepsis-3 onset time	-	13.0 (8.0–34.0)	-
<b>Ethnicity</b>			
White	40364 (63.6)	2087 (62.9)	38277 (63.7)
Black	5809 (9.2)	262 (7.9)	5547 (9.2)
Asian	721 (1.1)	42 (1.3)	679 (1.1)
Hispanic	630 (1.0)	32 (1.0)	598 (1.0)
Other/Unknown	14924 (23.5)	897 (27.0)	14027 (23.3)
<b>Admission Type</b>			
Medical	45009 (71.0)	2817 (84.8)	42192 (70.2)
Surgical	2239 (3.5)	45 (1.4)	2194 (3.7)
Other/Unknown	15200 (24.0)	458 (13.8)	14742 (24.5)

### 6.1.3.1. Feature Choice

To enable direct result comparisons with [38] benchmark, their feature set is adopted, which has been derived in collaboration with clinical experts and includes only widely available clinical markers. Each patient in the final cohort has 52 input-features, with four static variables (age, height, and weight at admission as well as sex) and 48 dynamic time-series variables. Dynamic variables combine seven vital signs and 39 laboratory tests, and two additional measurements (fraction of inspired oxygen and urine output). Table 1.11. provides a complete listing of all features with their value ranges, units of measurement and clinical descriptions. Target variables additionally include SOFA-score, a SI label, in contrast to YAIB where only the Sepsis-3 label is used (see Section 5.1).

### 6.1.3.2. Preprocessing

Data preprocessing involves three main steps: scaling, sampling, and imputation of features. These steps were again adopted from YAIB. All numerical features were standardized to zero mean and unit variance, while categorical and binary features remained left unchanged. To prevent data leakage, all normalization statistics were computed exclusively from the training split and applied to all partitions.

All features were uniformly resampled to an hourly basis with every trajectory padded to the maximum length of 169 hours, ensuring uniform processing lengths. Padded time-points were masked out for training and evaluation. Missing data points for dynamic variables were forward-filled using the last known value of the same ICU stay. For missing values without any prior measurement, the training cohort mean is used as fill value instead. Lastly the input data is augmented by a binary indicator that distinguishes between actual measurements and imputed values.

## 6.2. Implementation Details

Implementation of the LDM was done in the JAX [55] based Equinox framework [69]. Following sections present the implementation details for each module along with their respective parameter counts.

The infection indicator module  $f_{\theta_f}$  is a single GRU-cell with a hidden size of  $H_f = 32$ , followed by the down-projection layer, in total this adds up to 13,249 parameters (component-specific parameter counts are listed in Table 1.13. of the Appendix A.4).

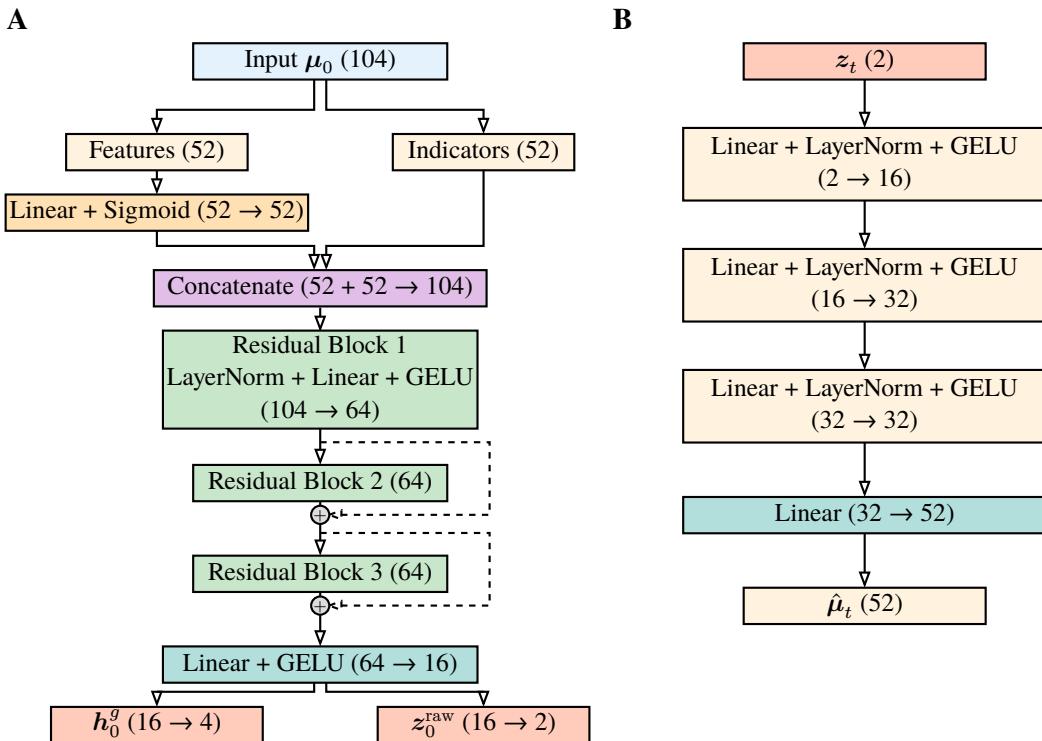
The latent encoder architecture  $g_{\theta_g}$  implements a gated attention mechanism with residual processing. Incoming samples of  $\mu_t$  are split into the actual features and the imputation indicator. A sigmoid-gate weights the features, effectively learning which features to emphasize. After gating, both halves are recombined and processed through a three-layer residual network, where each block applies layer-normalization [70]:

$$x' = \frac{x - \mathbb{E}[x]}{\text{Var}[x]} \cdot \gamma + \beta \quad (6.1)$$

where  $\gamma$  and  $\beta$  are learnable parameters, with the same dimensionality as  $x$ . After normalization, the input is processed by the Gaussian Error Linear Unit (GELU) activation [71], and a linear transformation with residual connections. With GELU:

$$\text{GELU}(x) = x\Phi(x) \quad (6.2)$$

where  $\Phi(x)$  is the cumulative distribution function for Gaussian distribution. The final hidden state passes through another GELU-activated layer before being projected into the two outputs  $\mathbf{h}_0$  and  $z_0^{\text{raw}}$ . The architecture is illustrated in panel A of Figure 6.16., in total the encoder has 19,350 parameter. The rollout module  $g_{\theta_g}$  performs latent space dynamics using a single GRU-cell, with a hidden size of  $H_g = 4$ , followed by the down projecting layer. This adds to 1,344 parameter. Choosing a small hidden dimension  $H_g$  simplifies post-hoc analysis of what information is carried inside hidden dimension, while still allowing for sufficient information preservation to steer the latent trajectory effectively. Though this analysis is out of the scope for this work and is not performed in this work.



**Figure 6.16.: A** Shows the initial latent encoder  $g_{\theta_e}$  architecture with feature gating and residual connections. Dashed arrows indicate residual skip connections. **B** Shows the decoder  $d_{\theta}^d$  architecture, reconstructing only the features but not the imputation indicators.

The decoder  $d_{\theta_d}$  is implemented as a four-layer feed-forward network that progressively up-samples from the latent representation back to the feature dimension, reconstructing only the features, not the imputation indicator. It uses GELU activations between layers to introduce non-linearity. The architecture is illustrated in panel **B** of Figure 6.16., in total the decoder has 3,524 parameters.

### 6.2.1. Training Details

The cohort was partitioned at the patient level using a stratified split with a 80/10/10 ratio for training, validation, and test sets respectively, yielding  $N = 50,740/6,343/6,342$  samples. Splitting was stratified by sepsis status to maintain the 5.2% prevalence ratio across all sets. The test set was reserved for final evaluation only and was not involved in hyperparameter tuning.

Hyperparameters were manually tuned rather than automatically optimized for two key reasons. First, the loss function jointly optimizes prediction accuracy and latent space interpretability. Automated hyperparameter optimization would require reducing these to a single scalar metric, which risks producing models with high predictive performance but degraded latent space interpretability, defeating a core purpose of the DNM integration. Second, exhaustive search over six loss components and 12 total tunable parameters would require hundreds of training runs, which is prohibitive for a proof-of-concept study. The primary goal is to demonstrate feasibility and interpretability rather than achieving state-of-the-art performance. Table 6.7. lists full hyperparameter specifications and initial values for learnable parameters.

Notably, the weight for the SOFA classification loss  $\mathcal{L}_{\text{SOFA}}$  is magnitudes larger than every other loss. Empirically, this has lead to the best results, and it is not uncommon for multitask optimizations that one loss outweighs others by multiple magnitudes [72]. In the end the size of the loss-gradients matter, not the raw numerical values.

All modules were jointly optimized using the AdamW-optimizer [73] with a mini-batch size of 128. The first epoch serves as a warm-up phase where the learning rate increases linearly from 0.0 to  $5 \times 10^{-5}$ . Subsequently, a constant learning rate is maintained for all remaining epochs. For the optimizer configuration, a weight decay of  $\lambda = 1 \times 10^{-1}$  is chosen, with momentum parameters  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ . DNM latent space was quantized to a  $60 \times 100$  grid over  $\beta \in [0.4\pi, 0.7\pi]$  and  $\sigma \in [0.0, 1.5]$ , with differentiable lookup using a  $3 \times 3$  neighborhood softmax interpolation ( $K = 9$ ).

Training was carried out on a consumer laptop GPU (NVIDIA RTX 500 Ada Generation with 4 GB of VRAM) on 32-bit floating-point precision. Training lasted a maximum of 1000 epochs, to prevent overfitting, early stopping was employed, where training stops after 30 consecutive epochs where both AUROC and AUPRC have not been improved. A typical training run to convergence required approximately 40 minutes, with early stopping occurring between epoch 100 to 400. To evaluate model performance against the held-out test set, the parameter configuration at the geometric mean between the epochs where AUROC and AUPRC peak is selected as the final model.

**Table 6.7.:** Hyperparameters and initial values for learnable parameters. Hyperparameters control training dynamics and loss weighting. Learnable parameters are initialized to the listed values and updated during training.

PARAMETER	VALUE	DESCRIPTION	REFERENCE
<b>Hyperparameter</b>			
$\lambda_{\text{sepsis}}$	600.0	Weight of $\mathcal{L}_{\text{sepsis}}$	Equation (5.40)
$\lambda_{\text{sofa}}$	$2 \times 10^3$	Weight of $\mathcal{L}_{\text{sofa}}$	Equation (5.40)
$\lambda_{\text{inf}}$	1.0	Weight of $\mathcal{L}_{\text{inf}}$	Equation (5.40)
$\lambda_{\text{spread}}$	$6 \times 10^{-3}$	Weight of $\mathcal{L}_{\text{spread}}$	Equation (5.40)
$\lambda_{\text{boundary}}$	30.0	Weight of $\mathcal{L}_{\text{boundary}}$	Equation (5.40)
$\lambda_{\text{dec}}$	5.0	Weight of $\mathcal{L}_{\text{dec}}$	Equation (5.40)
$\tau$	12	Radius of causal smoothing	Equation (5.36)
$k$	3	Side length of the latent-lookup kernel	Equation (5.29) Equation (5.31)
<b>Learnable Parameter</b>			
$d$	0.04	SOFA increase detection threshold	Equation (5.10)
$s$	50	SOFA increase detection sharpness	Equation (5.10)
$T_d$	0.05	Lookup interpolation temperature	Equation (5.29)
$\alpha$	0.7	Causal smoothing decay	Equation (5.36)

## 6.3. Results

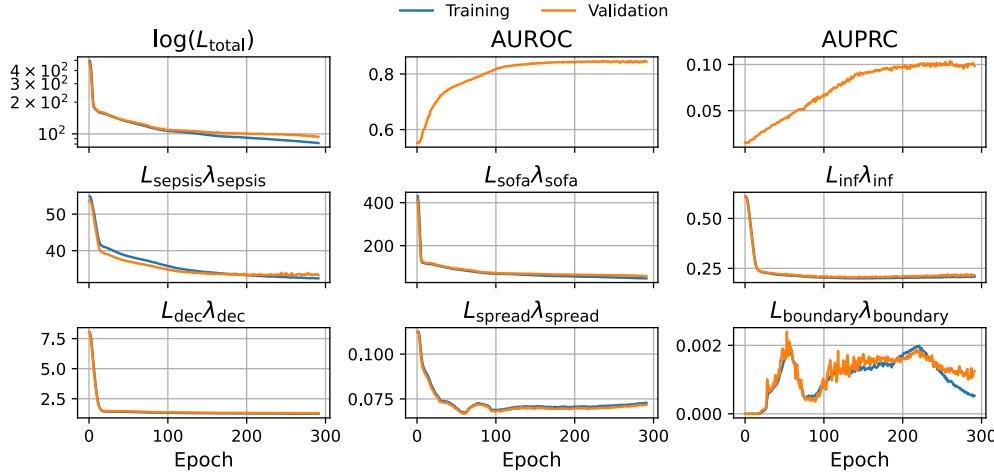
Following sections present the experimental results generated by training the LDM on the MIMIC-IV dataset. Details on the implementation and hyperparameters are specified in the previous two sections. Here, the training progression is discussed in Section 6.3.1, quantitative results presented in Section 6.3.2 followed by qualitative analyses in Section 6.3.3.

### 6.3.1. Training Progression

Figure 6.17. shows the progression of total loss  $\mathcal{L}_{\text{total}}$  and each loss component on both the training and validation set, as well as the metrics AUROC and AUPRC on the validation set. The total loss  $\mathcal{L}_{\text{total}}$  rapidly decreases in the first 10 epochs in both the training and validation set, thereafter, declining more gradually. For the majority of training time, the curves closely align, indicating stable multitask, showing only slight signs of overfitting towards the end. Total training loss decreases from 493 to 82, while the total validation loss reaches a slightly higher value of 94.

Analyzing component losses  $\mathcal{L}_{\text{sofa}}$  and  $\mathcal{L}_{\text{inf}}$  does not provide much further insight into the learning dynamics. Both validation curves decrease rapidly and stabilize afterwards, indicating the respective modules quickly learn robust representation, namely the alignment of the

latent space with organ dysfunction, as well as SI. The  $\mathcal{L}_{\text{inf}}$  is also indirectly influenced by the sepsis-sample upsampling, causing a slightly worse training loss. This behavior is also found in the primary sepsis loss  $\mathcal{L}_{\text{sepsis}}$ , sharing the same dynamics of strong early decrease, followed by a stabilization phase.



**Figure 6.17.:** Training and validation curves of the LDM showing the evolution of the total loss, task metrics (AUROC, AUPRC), and all individual loss components. The plots illustrate stable multi-objective convergence, early alignment of infection  $f_{\theta_f}$  and SOFA  $g_{\theta_g}$  submodules, and gradual refinement of the final sepsis risk prediction without signs of overfitting.

Decoder reconstruction loss  $\mathcal{L}_{\text{dec}}$  exhibits an overall healthy learning progression, again decreasing early and followed by a complete stabilization, indicating that it is able to fulfill main role as a structural regularizer of the latent space.

Spread loss  $\mathcal{L}_{\text{spread}}$  decreases gradually initially, preventing latent collapse, and later rises slightly, indicating a balanced trade-off between latent diversity and task alignment. Starting from 0.0, since no latent points are placed close to the edges of the parameter space,  $\mathcal{L}_{\text{boundary}}$  starts increasing gradually over training, as the model carefully learns to approach the space boundaries.

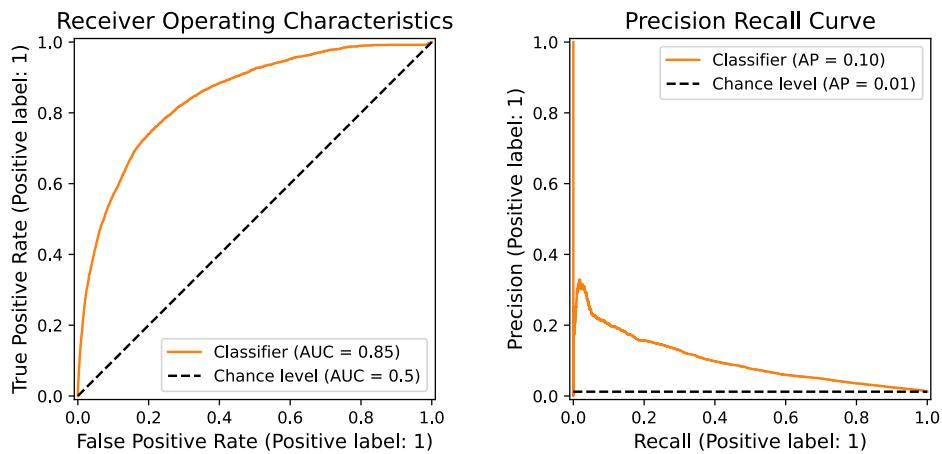
By optimizing these losses, the model improves at predicting the sepsis label from EHR histories. Quantitatively, validation AUROC rises from near random performance (0.5) to 0.847, with most gains occurring in the first third of training, followed by smaller but consistent improvements. Similarly, validation AUPRC increases monotonically, which is particularly relevant given the strong class imbalance. Absence of any significant late-stage decline in these metrics suggests that the model continues refining clinically meaningful discrimination rather than memorizing training data, though both slightly decrease after peaking.

### 6.3.2. Quantitative Results

This training run was early-stopped after 290 epochs, with 258 being the selected epoch, from validation AUROC peak at epoch 255 with value 0.847 and AUPRC peak at epoch 260 with value 0.104. Evaluating the model on the held-out test set, complete curves of the ROC and

precision-recall curves are shown in Figure 6.18. by sweeping the decision threshold  $\tau$  from Equation (5.41). Here, the ROC curve exhibits an expected shape, bending toward the top-left corner and away from the chance level curve shown as the dashed line. An AUROC of 0.84 tells us that the model will assign ground truth sepsis patients a higher risk  $\tilde{S}_t$  than non-septic patients in 84.69% of the time.

In contrast, the precision-recall curve shows higher precision at lower recall values (corresponding to lower threshold values  $\tau$ ), indicating the model struggles to produce confident positive predictions without predicting many false positives. On average, when model predicts a positive label, it is correct only 9.87% of the time, compared to a chance level of 1% determined by the prevalence of sepsis positive time window ( 1% of hours in the cohort with 5% septic patients).



**Figure 6.18.:** Receiver Operating Characteristics and precision-recall curves generated by sweeping the detection threshold  $\tau$  from Equation (5.41).

To put these results into perspective, Table 6.8. compares the LDM performance to baseline models trained in YAIB [38] on the same task, data-source, cohort definition and train-test split. The baseline models include include classical ML and DL approaches, with hyperparameters tuned via Bayesian optimization, 30 iterations for ML models, 50 for DL models. The LDM outperforms all baselines in both metrics.

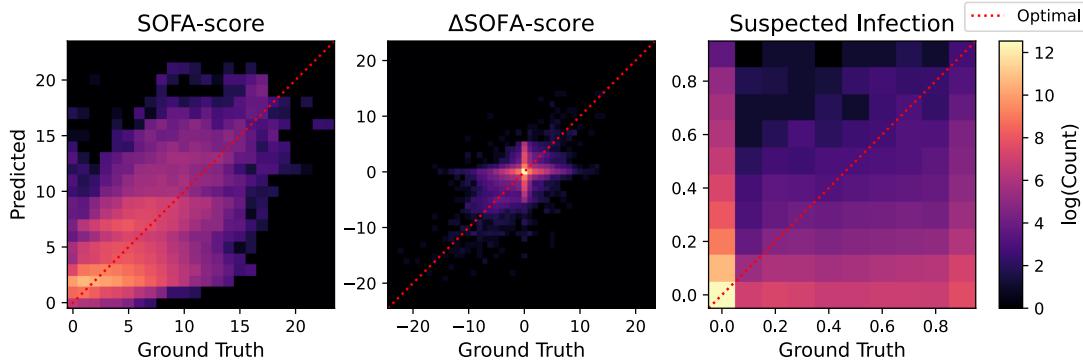
**Table 6.8.:** Performance Comparison against the mean performance of the baseline models trained in YAIB [38], in terms of AUROC  $\times 100$  ( $\uparrow$ , higher is better) and AUPRC  $\times 100$  ( $\uparrow$ ). Best performances in **bold**.

MODEL	AUROC	AUPRC
<b>YAIB</b>		
Regularized Logistic Regression	77.1	4.6
Light Gradient Boosting Machine	77.5	5.9
Gated Recurrent Unit	<b>83.6</b>	<b>9.1</b>
Long Short-Term Memory	82.0	8.0
Temporal Convolutional Network	82.7	8.8
Transformer	80.0	6.6
<b>This work</b>		
Dynamic Network Model	<b>84.69</b>	<b>9.87</b>

In Figure 6.19, sample-density plots compare predicted and ground-truth values, demonstrating that the model captures both the magnitude and temporal dynamics of organ dysfunction and infection. For each variable a diagonal line corresponds to optimal prediction performance. Predictions for the magnitude in SOFA-score follow the diagonal trend across the full severity range. This is indicating that the model preserves the ordinal structure of organ failure rather than collapsing toward the mean, though here it distributes most of the mass. While extreme SOFA values are slightly smoothed, the overall distribution is well reproduced, suggesting that the latent representation inside the DNM parameter space is able to capture clinically meaningful severity information.

For the change in SOFA score ( $\Delta$ SOFA, of consecutive time-steps), the distribution is strongly centered around zero for both ground truth and predictions, reflecting that most time-steps do not involve large acute changes. The model captures this concentration as well as the spread toward positive and negative deviations, indicating that it learns not only absolute severity but also the direction and magnitude of temporal deterioration or recovery. Yet there remains some weight where the ground truth increases but the model predicts a decrease, and vice versa.

Density values of prediction  $\tilde{I}_t$  vs. ground truth  $I_t$  show a mild separation between low and high infection probabilities. It is highly concentrated around non-infectious ground-truth values, but one can see a staircase pattern starting from correct non-infectious, leading up to correct infectious predictions. Together, these results indicate that the models internal representations align well with the clinical variables that define sepsis. Prediction densities are not shown for the sepsis label  $S_t$  since the strong imbalance renders this visualization uninformative even with log scaling.

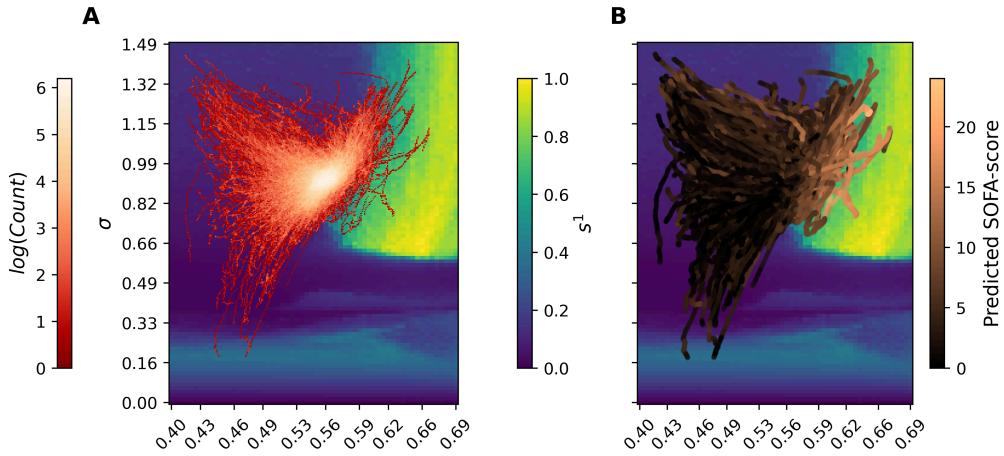


**Figure 6.19.:** Density plots comparing ground truth and predicted values for SOFA score, immediate SOFA change ( $\Delta$ SOFA), and SI. The model captures the overall SOFA severity distribution and its temporal changes, while infection predictions show reasonable separation between low- and high-probability states. Color indicates log sample density.

### 6.3.3. Qualitative Results

This section investigates, whether the models is able to create plausible patient trajectories inside DNM parameter space. First of all, in panel **A** of Figure 6.20. shows predicted sample density of  $z$  layered over the latent space, where both the smooth low-desynchronized between  $\beta$  values 0.4 and 0.5 and the more dynamic highly-desynchronized area between  $\beta$  values 0.4 and 1.0 are occupied. The lower-right part of the latent space is completely unused. The distribution strongly centers around the point  $z_{\text{center}} \approx (\beta = 0.58, \sigma = 0.98)$ , from there, individual trajectories spread out into all directions, mostly ending up in the low-desynchronized.

In panel **B**, the predicted latent positions colored by the ground truth SOFA-scores are layered over the latent space. Since brighter colors for predicted  $s^1(z)$  and ground truth correspond to higher desynchronization values, and therefore a more pathologic state of the organ system, ideally the gradients perfectly match. For the investigated model, this is not the case, but the general systematics align, where the majority of darker shaded trajectories overlap darker areas, and the same for brighter shades. This is indicating that the model is able to systematically use the DNM parameter space to express a patients physiological organ system state.



**Figure 6.20.**: **A** shows the distribution of predicted latent points  $z = (z_\beta, z_\sigma)$  over the latent space. The latent space is colored with the values of the normalized desynchronization metric  $s_z^1$ , where brighter values indicate larger desynchronization. The point distribution is colored by density, with brighter values having greater density. **B** shows the same latent space, but the overlay points are colored by the ground truth SOFA-score, here it is desired that the color gradients align.

Moving from the system wide behavior to three individual patient trajectories shown in Figure 6.21.. These hand-picked examples illustrate possible physiological progressions and how each is represented inside the DNM parameter space. The figure includes Patient 1 indicated by cyan, Patient 2 by purple and Patient 3 by pink colors. Their predicted and ground truth SOFA-score evolution is shown on the right side of the figure, with the left side showing corresponding latent trajectories colored by the ground truth values.

Patient 1 has at ICU admission  $t = 0$  a relatively low SOFA-score of 12, and the organ system deteriorates further and the score increases to value of 18 and settles down at a final value of 17. In the beginning, the prediction strongly underestimates the severity of Patient 1's condition by a value of 10, yet it picks up the trend of deterioration, but overshoots and arrives at a SOFA-score of 20. This progression is mirrors in the latent space, where the points clearly move from darker background-colors (more synchronization, better organ functionality) to brighter ones (less synchronization, worse organ functionality).

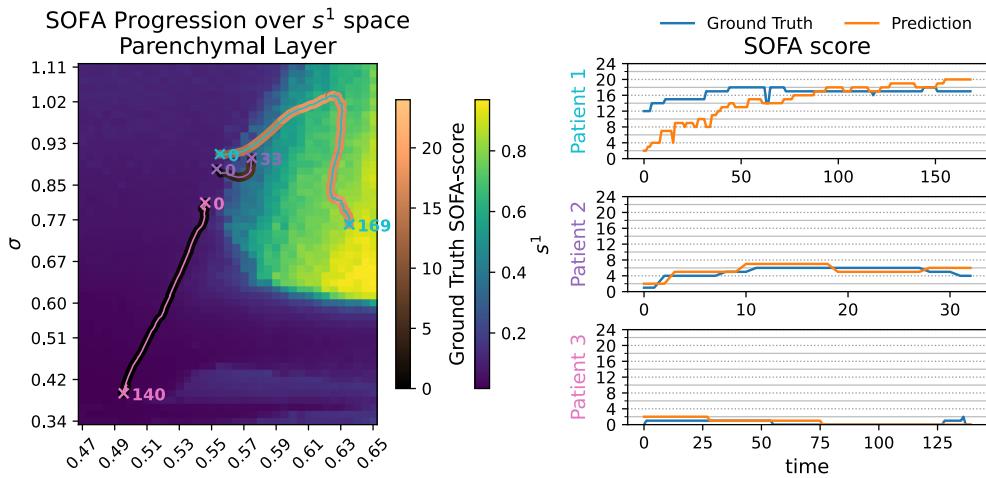
Patient 2 illustrates a recovery trajectory. Arriving with a SOFA score of 1, the patient experiences a brief deterioration to a score of 7 before recovering to a score of 2, approximately around 26 hours post-admission. The model tracks this directional change, though with imprecise magnitude estimates. In latent space, the trajectory initially moves toward the brighter right region, then rotates back into the darker zone as recovery progresses. Notably, while predicted SOFA scores do not fully align with ground truth values toward the end, the trajectory's rotational pattern clearly indicates recovery, the model captures the qualitative shift even when quantitative predictions deviate.

Similarly, Patient 3 arrives with an initial SOFA-score of 0, which increases shortly to 1 over two days and returns to the baseline value 0, after 6 days the SOFA score peaks for a few hours. The prediction does not detect the slight increase and stays at its initial value of 2 for the first day before decreasing to 1 and down to 0 after two more days. Initially, the latent

trajectory of Patient 3 moves parallel to the brighter region, followed by a turn away into the darker region, predicting a healthy progression, the small peak is not picked up anymore.

Overall these three exemplary patients show that the model can capture diverse physiological trajectories within the DNM parameter space. While the magnitude of predicted SOFA-scores sometimes deviates from ground truth, the model consistently captures directional trends, like deterioration, recovery, and stability, through meaningful movement in latent space. Correspondence between trajectory direction and organ dysfunction severity suggests the model has learned clinically relevant representations that map physiological states to interpretable DNM parameters.

However, it should be noted that these hand-picked examples demonstrate typical model behavior, though performance varies across patients. Some trajectories show better alignment with clinical progression, while others exhibit larger deviations from ground truth, this is somewhat reflected by the deviations in Figure 6.19. and heterogeneity of trajectories in panel **B** in Figure 6.20..



**Figure 6.21.:** Left: Three patient trajectories plotted in the latent DNM parameter space ( $\beta, \sigma$ ), colored by ground truth SOFA scores with timestamps marking trajectory progression. The background heatmap shows  $s^1$  values (right colorbar), with darker regions indicating lower desynchronization (better organ function) and brighter regions higher desynchronization (worse organ function). Right: Time series comparing predicted (orange) versus ground truth (blue) SOFA scores for each patient. Time annotations on trajectories indicate beginning and end of temporal progression. While prediction magnitudes occasionally diverge from ground truth, the model consistently captures directional trends through systematic latent space navigation.



# 7

## Discussion

This work introduced the Latent Dynamics Model (LDM), a novel architecture that embeds the Dynamic Network Model (DNM) as an interpretable latent space for short-term sepsis prediction from Electronic Health Record (EHR). By combining physics-inspired structure with data-driven learning, this hybrid approach demonstrates competitive predictive performance while offering enhanced interpretability compared to conventional models. This chapter discusses the key findings in relation to the research questions posed in Section 4.4, addresses limitations of the current implementation, and outlines directions for future research.

### 7.1. Addressing the Research Questions

This work was motivated by two central questions regarding the clinical utility of the DNM:

***Usability of the DNM:*** *How and to what extent can ML-determined trajectories of the DNM be used for detection and prediction, especially of critical infection states?*

***Comparison with data-based approaches:*** *How can the model-based predictions be compared with those of purely data-based approaches in terms of predictive power?*

Experimental results provide encouraging answers to both questions. Regarding usability, it was demonstrated that DNM parameters ( $\beta, \sigma$ ) can be inferred from clinical EHR time-series data, and that trajectories through this parameter space capture clinically meaningful disease progression. Systematic correspondence between latent positions and clinical severity suggests that DNM parameters learned from data do correlate with disease trajectory, support the physiological relevance of the model.

Regarding comparison with data-driven approaches, LDM achieved an AUROC of 84.69% and AUPRC of 9.87%, slightly outperforming all baseline models from the YAIB benchmark, including the best-performing standard GRU (83.6% AUROC, and 9.1% AUPRC). Critically, this performance was achieved while providing medically interpretable intermediate representations, whereas baseline models operated as complete black boxes. This suggests that embedding DNM structure does not sacrifice predictive power and may actually provide useful inductive biases that facilitate learning clinically relevant patterns.

## 7.2. Key Strengths

Perhaps the most significant finding is that incorporating physiologically-motivated structure does not compromise predictive accuracy. Compared to most baseline models which provide only a single risk-score, whereas LDM provides multiple clinically interpretable indicators, namely infection likelihood  $\tilde{I}_t$ , organ desynchronization  $s_t^1(\hat{z})$ , acute deterioration risk  $\tilde{A}_t$ , and overall sepsis risk  $\tilde{S}_t$ . This richer output enables clinicians to understand **why** a patient is flagged as high-risk, supporting more informed decision-making. Importantly, the LDM interpretability is not post-hoc rationalization but structurally embedded in the model architecture. Furthermore, traditionally, predictive models undergo extensive hyperparameter optimization, the LDM parameters were manually tuned with emphasis on maintaining latent space interpretability rather than maximizing performance metrics alone.

As discussed in Section 3, explainability in current data-driven sepsis prediction systems predominantly relies on Shapley-value analyses, deriving importance factors of single input features or feature interactions [35], [36]. While valuable, such approaches explain which features influenced a prediction without revealing how those features interact dynamically to produce physiological states. In contrast, DNM-based trajectories show temporal evolution through a space with direct physiological interpretation, while still allowing conventional feature-importance analyses.

This interpretability operates at multiple levels. At the population level, density plots (Figure 6.20.) reveal that the model systematically organizes patients according to severity. At an individual patient level, trajectory shapes encode temporal dynamics, sharp directional changes correspond to acute clinical events, while gradual curves indicate slower progression or recovery (Figure 6.21.). This multi-scale interpretability aligns with clinician preferences identified in [24], where survey participants emphasized that “the trend of a patient’s trajectory itself should be the prediction target” and expressed preference for “trajectories over plain binary event predictions.”

With roughly 21,000 parameters, the entire LDM is a relatively small network. Additionally, due to its modularity into functional clear roles, i.e.  $f_{\theta_f}$ ,  $g_{\theta_g}$  and  $d_{\theta_d}$ , it provides the necessary flexibility to improve on individual aspects of the system.

## 7.3. Limitations and Challenges

While the experimental results are encouraging, several important limitations warrant careful consideration. Preliminary experiments revealed notable sensitivity to random seed initialization. Different random seeds produced models with qualitatively different latent space organizations. While final predictive performance remained relatively stable ( $\pm 2 - 3\%$  AUROC), the specific geometric arrangement of patients in parameter space varied considerably. This sensitivity likely stems from the multi-objective loss function, which creates a complex optimization landscape with multiple local minima, and globally not with a single optimal solution, rather a Pareto frontier of trade-offs.

---

Additionally, there is most likely no ground truth mapping from a patient state, represented by the EHR, to the low dimensional parameter space of the DNM. This ambiguity offers infinitely many mappings, while others might be more plausible than others, each training run converges most likely to a different mapping depending on the random seed. Ultimately, when interpreting results, the variability in solutions should be acknowledged rather than treating a single trained model as definitive. Ensemble methods aggregating predictions from multiple initializations could improve robustness while quantifying this uncertainty.

As discussed in Section 4.4, the DNM faces inherent limitations. Parameters like  $\beta$  (biological age) and  $\sigma$  (interaction strength between organ and immune system) do not correlate to any directly observable physiological quantity. Furthermore, the fully connected topology may not reflect actual organ interaction patterns, and individual oscillators do not correspond to specific biological processes. Currently, the LDM does not directly implement the DNM dynamical system, since it does not solve the coupled differential equations at each time step. Instead, it learns to position patients in a parameter space abstracted from the DNM, with the hope (encouraged by loss functions) that these positions correlate meaningfully with true physiological states. It is learning a projection from high-dimensional EHR data into a two-dimensional space. The desynchronization metric  $s^1(z)$  computed from these projected positions may correlate with organ dysfunction without reflecting actual organ-level dynamics. Experimental results provide some reassurance, the systematic correlation between latent position and SOFA scores, the meaningful trajectory patterns in Figure 6.21., and the competitive predictive performance all suggest the model has learned clinically relevant structure. However, direct mechanistic interpretability cannot be claimed, a more honest characterization might be “physiologically-motivated dimensionality reduction”.

Right now, the LDM produces risk-scores  $\tilde{S}_t$  indicating sepsis likelihood, but they have not been calibrated to represent true probabilities of sepsis onset. This means, the prediction has to move from plain distinction septic vs. non-septic to sensitive estimates, how critical patient states are compared to others. For clinical deployment, well-calibrated probabilities would be essential, which could be achieved with post-training calibration, like simple Platt- or Temperature scaling or more involved calibration techniques [74]. As noted in Section 2.3, traditional sepsis screening relies on reactive clinical scores like SOFA, while automated prediction systems aim to identify patients before organ failure develops. However, as highlighted by the meta-analysis in [20], many alert systems fail to improve patient outcomes, often due to alert fatigue from excessive false positives.

## 7.4. Future Directions

All experiments used MIMIC-IV exclusively. While this enables direct comparison with YAIB benchmarks, it limits generalizability claims. As noted in Section 3, applications trained on MIMIC databases “are known to have limited generalization to other datasources and real-world settings” despite being “the default open-data resource for developing sepsis prediction systems.” External validation on independent datasets is essential to assess whether learned representations transfer across settings, but are out of this scope for this

proof-of-concept thesis. Some performance degradation on external data is expected, but the central question is whether the DNM-structured latent space provides robust representations. If the latent space captures fundamental physiological principles rather than dataset-specific patterns, these representations should be more transferable.

Current evaluation focused primarily on predictive metrics and trajectory visualization, more rigorous analysis of latent space structure and model behavior could provide deeper insights. For example systematically assess whether  $d_{\theta_d}$  successfully regularizes the latent space, in a way that disentangles clinical concepts in a meaningful way. If it does not, can it be achieved? Could this information used to deduce practical patient treatment?

Furthermore, the encoder  $g_{\theta_g^e}$  uses sigmoid gates to weight input features. Analyzing learned gate weights could reveal, which features the model considers most informative for positioning patients in DNM space, whether gating patterns differ between septic and non-septic patients or if gates encode any known clinical knowledge. Lastly quantitative analysis of latent trajectory curvature, velocity, and acceleration could give insights if these geometric properties correlate with clinical severity or outcomes. Such analyses would either strengthen confidence in the DNMs clinical validity or reveal specific weaknesses requiring architectural modification.

Current implementation uses a differential lookup methodology to retrieve the DNM synchronization metrics. A more principled approach would directly integrate the DNM system of differential equations: At each time step, given latent position  $\mathbf{z}_t = (\beta_t, \sigma_t)$ , solve this system numerically to obtain  $s^1(\mathbf{z}_t)$ . Modern differentiable ODE-solvers, such as diffraz [56], enable backpropagation through the integration process, making this approach trainable end-to-end. This approach would come with a tremendous increase in computational cost, but is the most promising path toward actual mechanistic and more nuanced interpretability.

Current work focused on online prediction, where risk estimates update continuously as new measurements arrive. As noted in Section 3, offline prediction tasks, meaning predicting sepsis risk at a fixed observation time for a specified future horizon  $T$ , are also clinically relevant. Because of the LDMs modularity, an extension to handle offline predictions is straightforward and would require minimal changes. This would test whether learned latent dynamics capture sufficient structure to extrapolate forward in time, and whether DNM-based trajectories provide better forecasting than purely data-driven alternatives. Training a model on both online- and offline-prediction might offer cross-benefits for both tasks.

While DNM provides a physiologically motivated latent space, other spaces might also provide interpretable structures, as long as there are regions that could be assigned to healthy and septic patients and some interpretation behind the space dimensions. Even not biologically motivated but artificially created spaces could be used for the systematic comparison, revealing which structures best capture sepsis pathophysiology, or if its solely the model built around it that drives prediction performance.

Lastly, some general concerns, not specific to the LDM. Current sepsis prediction implementation treat patients as passively observed systems, ignoring medical interventions. Incorporating treatment variables could improve predictions and enable counterfactual reasoning about intervention timing, though this raises its own methodological challenges.

Furthermore, hourly resampling with forward-fill imputation may obscure rapid deterioration occurring between measurements. More sophisticated approaches like irregular time series methods could better capture high-frequency dynamics.



# Appendix

A Appendix .....	73
A.1 SOFA - Details .....	73
A.2 DNM as Lie Formulation .....	74
A.3 Input Concepts .....	74
A.4 Latent Dynamic Model Architecture Parameters .....	76

## A Appendix

### A.1 SOFA - Details

**Table 1.9.:** TODO

CATEGORY	INDICATOR	1	2	3	4
Respiration	PaO <sub>2</sub> /FiO <sub>2</sub> [mmHg]	< 400	< 300	< 200	< 100
	Mechanical Ventilation			yes	yes
Coagulation	Platelets [ $\times \frac{10^3}{\text{mm}^3}$ ]	< 150	< 100	< 50	< 20
Liver	Bilirubin [ $\frac{\text{mg}}{\text{dl}}$ ]	1.2-1.9	2.0-5.9	6.0-11.9	> 12.0
Cardiovascular <sup>4</sup>	MAP [mmHg]	< 70			
	or Dopamine		$\leq 5$	> 5	> 15
	or Dobutamine		any dose		
	or Epinephrine			$\leq 0.1$	> 0.1
	or Noepinephrine			$\leq 0.1$	> 0.1
Central Nervous System	Glasgow Coma Score	13-14	10-12	6-9	< 6
Renal	Creatinine [ $\frac{\text{mg}}{\text{dl}}$ ]	1.2-1.9	2.0-3.4	3.5-4.9	> 5.0
	or Urine Output [ $\frac{\text{ml}}{\text{day}}$ ]			< 500	< 200

<sup>4</sup>Adrenergic agents administered for at least 1h (doses given are in [ $\mu\text{g}/\text{kg} \cdot \text{min}$ ])

## A.2 DNM as Lie Formulation

## A.3 Input Concepts

**Table 1.10.:** Static input features for the prediction task

RICU - NAME	UNIT	MIN	MAX	DESCRIPTION
age	Years	0	-	Age at hospital admission
sex	-	-	-	Female Sex
height	kg	0	-	Patient height
weight	cm	0	-	Patient weight

**Table 1.11.:** Dynamic input features for the prediction task.

RICU - NAME	UNIT	MIN	MAX	DESCRIPTION
alb	g/dL	0	6	albumin
alp	IU/L, U/l	0	-	alkaline phosphatase
alt	IU/L, U/l	0	-	alanine aminotransferase
ast	IU/L, U/l	0	-	aspartate aminotransferase
be	mEq/L, mmol/l	-25	25	base excess
bicar	mEq/L, mmol/l	5	50	bicarbonate
bili	mg/dL	0	100	total bilirubin
bili_dir	mg/dL	0	50	bilirubin direct
bnd	%	-	-	band form neutrophils
bun	mg/dL	0	200	blood urea nitrogen
ca	mg/dL	4	20	calcium
cai	mmol/L	0.5	2	calcium ionized
ck	IU/L, U/l	0	-	creatine kinase
ckmb	ng/mL	0	-	creatine kinase MB
cl	mEq/L, mmol/l	80	130	chloride
crea	mg/dL	0	15	creatinine
crp	mg/L	0	-	C-reactive protein
dbp	mmHg, mm Hg	0	200	diastolic blood pressure
fgn	mg/dL	0	1500	fibrinogen
fio2	%	21	100	fraction of inspired oxygen
glu	mg/dL	0	1000	glucose
hgb	g/dL	4	18	hemoglobin
hr	bpm, /min	0	300	heart rate

inr_pt	-	-	-	prothrombin time/international normalized ratio
k	mEq/L, mmol/l	0	10	potassium
lact	mmol/L	0	50	lactate
lymph	%	0	100	lymphocytes
map	mmHg, mm Hg	0	250	mean arterial pressure
mch	pg	0	-	mean cell hemoglobin
mchc	%	20	50	mean corpuscular hemoglobin concentration
mcv	fL	50	150	mean corpuscular volume
methb	%	0	100	methemoglobin
mg	mg/dL	0.5	5	magnesium
na	mEq/L, mmol/l	110	165	sodium
neut	%	0	100	neutrophils
o2sat	%, % Sat.	50	100	oxygen saturation
pco2	mmHg, mm Hg	10	150	CO2 partial pressure
ph	-	6.8	8	pH of blood
phos	mg/dL	0	40	phosphate
plt	K/uL, G/l	5	1200	platelet count
po2	mmHg, mm Hg	40	600	O2 partial pressure
ptt	sec	0	-	partial thromboplastin time
resp	insp/min, /min	0	120	respiratory rate
sbp	mmHg, mm Hg	0	300	systolic blood pressure
temp	C, °C	32	42	temperature
tnt	ng/mL	0	-	troponin t
urine	mL	0	2000	urine output
wbc	K/uL, G/l	0	-	white blood cell count

## A.4 Latent Dynamic Model Architecture Parameters

**Table 1.12.:** Detailed parameter count of every module

NAME	SHAPE	COUNT
<b>Infection Module</b>		
GRU-Cell weight_ih	(96, 104)	9,984
GRU-Cell weight_hh	(96, 32)	3,072
GRU-Cell bias	(96,)	96
GRU-Cell bias_n	(32,)	32
h0	(32,)	32
Projection weight	(1, 32)	32
Projection bias	(1,)	1
<b>Total:</b>		<b>13,249</b>
<b>Decoder Module</b>		
linear1 weight	(16, 2)	32
linear1 bias	(16,)	16
norm1 weight	(16,)	16
norm1 bias	(16,)	16
linear2 weight	(32, 16)	512
linear2 bias	(32,)	32
norm2 weight	(32,)	32
norm2 bias	(32,)	32
linear3 weight	(32, 32)	1,024
linear3 bias	(32,)	32
norm3 weight	(32,)	32
norm3 bias	(32,)	32
linear4 weight	(52, 32)	1,664
linear4 bias	(52,)	52
<b>Total:</b>		<b>3,524</b>
<b>General</b>		
lookup_temp	(1,)	1
causal_decay	(1,)	1
detection_threshold	(1,)	1
detection_sharpness	(1,)	1
<b>Total:</b>		<b>4</b>

NAME	SHAPE	COUNT
<b>SOFA Module (Encoder)</b>		
gating weight	(52, 52)	2,704
norm1 weight	(104,)	104
norm1 bias	(104,)	104
linear1 weight	(64, 104)	6,656
linear1 bias	(64,)	64
norm2 weight	(64,)	64
norm2 bias	(64,)	64
linear2 weight	(64, 64)	4,096
linear2 bias	(64,)	64
norm3 weight	(64,)	64
norm3 bias	(64,)	64
linear3 weight	(64, 64)	4,096
linear3 bias	(64,)	64
linear4 weight	(16, 64)	1,024
linear4 bias	(16,)	16
z weight	(2, 16)	32
z bias	(2,)	2
h weight	(4, 16)	64
h bias	(4,)	4
<b>Total:</b>		<b>19,350</b>
<b>SOFA Module (RNN)</b>		
GRU-Cell weight_ih	(12, 106)	1,272
GRU-Cell weight_hh	(12, 4)	48
GRU-Cell bias	(12,)	12
GRU-Cell bias_n	(4,)	4
Proj weight	(2, 4)	8
<b>Total:</b>		<b>1,344</b>

## Bibliography

- [1] K. E. Rudd and 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, 2020, doi: 10.1016/S0140-6736(19)32989-7.
- [2] E. C. van der Slikke, A. Y. An, R. E. Hancock, and H. R. Bouma, “Exploring the pathophysiology of post-sepsis syndrome to identify therapeutic opportunities,” *EBioMedicine*, vol. 61, p. 103044, 2020, doi: 10.1016/j.ebiom.2020.103044.
- [3] C. Fleischmann-Struzek, D. Schwarzkopf, and K. Reinhart, “Inzidenz der Sepsis in Deutschland und weltweit: Aktueller Wissensstand und Limitationen der Erhebung in Abrechnungsdaten,” *Medizinische Klinik - Intensivmedizin und Notfallmedizin*, vol. 117, no. 4, pp. 264–268, May 2022, doi: 10.1007/s00063-021-00777-5.
- [4] M. Singer *et al.*, “The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3),” *JAMA*, vol. 315, no. 8, pp. 801–810, Feb. 2016, doi: 10.1001/jama.2016.0287.
- [5] Y. Chen and L. Tang, “The Crosstalk between Parenchymal Cells and Macrophages: A Keeper of Tissue Homeostasis,” *Frontiers in Immunology*, vol. 13, p. 1050188, 2022, doi: 10.3389/fimmu.2022.1050188.
- [6] A. M. Honan and Z. Chen, “Stromal Cells Underlining the Paths From Autoimmunity, Inflammation to Cancer With Roles Beyond Structural and Nutritional Support,” *Frontiers in Cell and Developmental Biology*, 2021, doi: 10.3389/fcell.2021.658984.
- [7] D. H. P. Directors, “Overview of Surface Epithelium.” [Online]. Available: <https://digitalhistology.org/tissues/epithelium/lining-and-covering/overview/overview-1/>
- [8] D. H. P. Directors, “Stroma and Parenchyma.” [Online]. Available: <https://digitalhistology.org/organs-systems/general-concepts/stroma-and-parenchyma/stroma-and-parenchyma-1/>
- [9] S. Fischer and E. Deindl, “State of the Art of Innate Immunity - An Overview,” *Cells*, vol. 11, p. 2705, 2022, doi: 10.3390/cells11172705.
- [10] D. Jarczak, S. Kluge, and A. Nierhaus, “Sepsis—Pathophysiology and Therapeutic Concepts,” *Frontiers in Medicine*, 2021, doi: 10.3389/fmed.2021.628302.
- [11] J. Zhang and J. An, “Cytokines, inflammation, and pain,” *Int Anesthesiol Clin*, vol. 45, no. 2, pp. 27–37, 2007, doi: 10.1097/AIA.0b013e318034194e.
- [12] R. V. House and J. Descotes, Eds., *Cytokines in Human Health: Immunotoxicology, Pathology, and Therapeutic Applications*, 1st ed. in Methods in Pharmacology and Toxicology. Totowa, N.J.: Humana Press, 2007.

- [13] D. Jarczak and A. Nierhaus, “Cytokine Storm—Definition, Causes, and Implications,” *International Journal of Molecular Sciences*, vol. 23, no. 19, 2022, doi: 10.3390/ijms231911740.
- [14] J. Lamsfus-Prieto, R. de Castro-Fernández, A. Hernández-García, and G. Marcano-Rodriguez, “Prognostic value of gasometric parameters of carbon dioxide in resuscitation of septic patients. A bibliography review,” *Revista Española de Anestesiología y Reanimación (English Edition)*, vol. 63, no. 4, pp. 220–230, 2016, doi: <https://doi.org/10.1016/j.redare.2015.11.003>.
- [15] J. L. Vincent *et al.*, “The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine,” *Intensive care medicine*, vol. 22, no. 7, pp. 707–710, 1996.
- [16] A. E. W. Johnson *et al.*, “A Comparative Analysis of Sepsis Identification Methods in an Electronic Database,” *Critical Care Medicine*, vol. 46, no. 4, pp. 494–499, Apr. 2018, doi: 10.1097/CCM.0000000000002965.
- [17] N. Bennett, D. Plečko, and I.-F. Ukor, “Sepsis 3 label — sep3.” 2025.
- [18] R. Moreno *et al.*, “The Sequential Organ Failure Assessment (SOFA) Score: has the time come for an update?,” *Critical Care*, vol. 27, no. 1, p. 15, Jan. 2023, doi: 10.1186/s13054-022-04290-9.
- [19] L. La Via *et al.*, “The Global Burden of Sepsis and Septic Shock,” *Epidemiologia*, vol. 5, no. 3, pp. 456–478, 2024, doi: 10.3390/epidemiologia5030032.
- [20] S. Ein Alshaeba, E. A. Marhoffer, J. L. Holleck, J. Theise-Toupal, A. A. Grimshaw, and C. G. Gunderson, “The Effect of Early Warning Systems for Sepsis on Mortality: A Systematic Review and Meta-analysis,” *Journal of General Internal Medicine (J GEN INTERN MED)*, vol. 40, pp. 3463–3468, 2025, doi: 10.1007/s11606-025-09569-5.
- [21] C. W. Seymour *et al.*, “Time to Treatment and Mortality during Mandated Emergency Care for Sepsis,” *The New England Journal of Medicine*, vol. 376, pp. 2235–2244, June 2017, doi: 10.1056/NEJMoa1703058.
- [22] G. A. Westphal *et al.*, “Early detection strategy and mortality reduction in severe sepsis,” *Revista Brasileira de Terapia Intensiva (Rev Bras Ter Intensiva)*, vol. 21, no. 2, pp. 113–123, June 2009.
- [23] G. P. Dobson, H. L. Letson, and J. L. Morris, “Revolution in sepsis: a symptoms-based to a systems-based approach?,” *Journal of Biomedical Science (J Biomed Sci)*, vol. 31, no. 1, p. 57, 2024, doi: 10.1186/s12929-024-01043-4.
- [24] B. Eini-Porat, O. Amir, D. Eytan, and U. Shalit, “Tell me something interesting: Clinical utility of machine learning prediction models in the ICU,” *Journal of Biomedical Informatics*, vol. 132, p. 104107, 2022, doi: 10.1016/j.jbi.2022.104107.
- [25] A. Schuurman, P. Sloot, W. Wiersinga, and et al., “Embracing complexity in sepsis,” *Crit Care*, vol. 27, no. 1, p. 102, 2023, doi: 10.1186/s13054-023-04374-0.
- [26] F. J. A. Relouw, M. Kox, H. R. Taal, B. C. P. Koch, M. W. J. Prins, and N. A. W. van Riel, “Mathematical model of the inflammatory response to acute and prolonged lipopolysaccharide exposure in humans,” *npj Systems Biology and Applications*, vol. 10, no. 1, p. 146, Dec. 2024, doi: 10.1038/s41540-024-00473-y.
- [27] G. An and C. Cockrell, “A design specification for Critical Illness Digital Twins to cure sepsis: responding to the National Academies of Sciences, Engineering and Medicine Report: Foundational Research Gaps and Future Directions for Digital Twins.” [Online]. Available: <https://arxiv.org/abs/2405.05301>
- [28] C. Cockrell, S. Schobel-McHugh, F. Lisboa, Y. Vodovotz, and G. An, “Generating synthetic data with a mechanism-based Critical Illness Digital Twin: Demonstration for Post Traumatic Acute Respiratory Distress Syndrome,” *bioRxiv*, 2023, doi: 10.1101/2022.11.22.517524.

- [29] M. McDaniel, J. M. Keller, S. White, and A. Baird, “A Whole-Body Mathematical Model of Sepsis Progression and Treatment Designed in the BioGears Physiology Engine,” *Frontiers in Physiology*, vol. 10, p. 1321, Oct. 2019, doi: 10.3389/fphys.2019.01321.
- [30] S. Bomrah *et al.*, “A scoping review of machine learning for sepsis prediction- feature engineering strategies and model performance: a step towards explainability,” *Critical Care*, vol. 28, no. 1, p. 180, May 2024, doi: 10.1186/s13054-024-04948-6.
- [31] M. Moor, B. Rieck, M. Horn, C. R. Jutzeler, and K. Borgwardt, “Early Prediction of Sepsis in the ICU Using Machine Learning: A Systematic Review,” *Frontiers in Medicine*, 2021, doi: 10.3389/fmed.2021.607952.
- [32] M. Y. Yadgarov *et al.*, “Early detection of sepsis using machine learning algorithms: a systematic review and network meta-analysis,” *Frontiers in Medicine*, 2024, doi: 10.3389/fmed.2024.1491358.
- [33] Y. Gao, C. Wang, J. Shen, Z. Wang, Y. Liu, and Y. Chai, “Systematic review and network meta-analysis of machine learning algorithms in sepsis prediction,” *Expert Systems with Applications*, vol. 245, p. 122982, 2024, doi: <https://doi.org/10.1016/j.eswa.2023.122982>.
- [34] S. P. A. and S. B., “Analysis of machine learning and deep learning prediction models for sepsis and neonatal sepsis: A systematic review,” *ICT Express*, vol. 9, no. 6, pp. 1215–1225, 2023, doi: <https://doi.org/10.1016/j.icte.2023.07.007>.
- [35] C. Stylianides *et al.*, “AI Advances in ICU with an Emphasis on Sepsis Prediction: An Overview,” *Machine Learning and Knowledge Extraction*, vol. 7, no. 1, 2025, doi: 10.3390/make7010006.
- [36] M. Sundararajan and A. Najmi, “The many Shapley values for model explanation,” in *Proceedings of the 37th International Conference on Machine Learning*, in ICML'20. JMLR.org, 2020.
- [37] A. E. W. Johnson and others, “MIMIC-IV, a freely accessible electronic health record dataset,” *Scientific Data*, vol. 10, no. 1, Jan. 2023, doi: 10.1038/s41597-022-01899-x.
- [38] R. van de Water, H. N. A. Schmidt, P. Elbers, P. Thoral, B. Arnrich, and P. Rockenschaub, “Yet Another ICU Benchmark: A Flexible Multi-Center Framework for Clinical ML,” in *The Twelfth International Conference on Learning Representations*, Oct. 2024.
- [39] R. E. Callard, A. J. T. George, and J. Stark, “Cytokines, chaos, and complexity,” *Immunity*, vol. 11, no. 5, pp. 507–513, Nov. 1999, doi: 10.1016/s1074-7613(00)80125-9.
- [40] P. C. Ivanov, “The New Field of Network Physiology: Building the Human Physiolome,” *Frontiers in Network Physiology*, 2021, doi: 10.3389/fnetp.2021.711778.
- [41] A. C. Guyton, T. G. Coleman, and H. J. Granger, “Circulation: overall regulation,” *Annu Rev Physiol*, vol. 34, pp. 13–46, 1972, doi: 10.1146/annurev.ph.34.030172.000305.
- [42] R. P. Bartsch, A. Y. Schumann, J. W. Kantelhardt, T. Penzel, and P. C. Ivanov, “Phase transitions in physiologic coupling,” *Proc Natl Acad Sci U S A*, vol. 109, no. 26, pp. 10181–10186, June 2012, doi: 10.1073/pnas.1204568109.
- [43] K. Lehnertz, T. Rings, and T. Bröhl, “Time in Brain: How Biological Rhythms Impact on EEG Signals and on EEG-Derived Brain Networks,” *Front. Netw. Physiol.*, vol. 1, p. 755016, Sept. 2021, doi: 10.3389/fnetp.2021.755016.
- [44] M. Madadi Asl, A.-H. Vahabie, A. Valizadeh, and P. A. Tass, “Spike-Timing-Dependent Plasticity Mediated by Dopamine and its Role in Parkinson’s Disease Pathophysiology,” *Frontiers in Network Physiology*, 2022, doi: 10.3389/fnetp.2022.817524.
- [45] N. Sinha, R. B. Joshi, M. R. S. Sandhu, T. I. Netoff, H. P. Zaveri, and K. Lehnertz, “Perspectives on Understanding Aberrant Brain Networks in Epilepsy,” *Frontiers in Network Physiology*, 2022, doi: 10.3389/fnetp.2022.868092.

- [46] J. Sawicki, R. Berner, T. Löser, and E. Schöll, “Modeling Tumor Disease and Sepsis by Networks of Adaptively Coupled Phase Oscillators,” *Frontiers in Network Physiology*, vol. 1, 2022, doi: 10.3389/fnetp.2021.730385.
- [47] R. Berner, J. Sawicki, M. Thiele, T. Löser, and E. Schöll, “Critical Parameters in Dynamic Network Modeling of Sepsis,” *Frontiers in Network Physiology*, vol. 2, 2022, doi: 10.3389/fnetp.2022.904480.
- [48] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Sciences*, vol. 12. Cambridge University Press, 2001.
- [49] Y. Kuramoto, *Chemical Oscillations, Waves, and Turbulence*, vol. 19. Berlin, Heidelberg: Springer-Verlag, 1984.
- [50] J. A. Acebrón, L. L. Bonilla, C. J. Pérez Vicente, F. Ritort, and R. Spigler, “The Kuramoto model: A simple paradigm for synchronization phenomena,” *Reviews of Modern Physics*, vol. 77, no. 1, pp. 137–185, 2005.
- [51] H. Sakaguchi and Y. Kuramoto, “A Soluble Active Rotator Model Showing Phase Transitions via Mutual Entrainment,” *Progress of Theoretical Physics*, vol. 76, no. 3, pp. 576–581, 1986, doi: 10.1143/PTP.76.576.
- [52] R. Berner, J. Sawicki, and E. Schöll, “Birth and Stabilization of Phase Clusters by Multiplexing of Adaptive Networks,” *Phys. Rev. Lett.*, vol. 124, no. 8, p. 88301, Feb. 2020, doi: 10.1103/PhysRevLett.124.088301.
- [53] B. Jüttner and E. A. Martens, “Complex dynamics in adaptive phase oscillator networks,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 33, no. 5, May 2023, doi: 10.1063/5.0133190.
- [54] R. Berner, J. Fialkowski, D. Kasatkin, V. Nekorkin, S. Yanchuk, and E. Schöll, “Hierarchical frequency clusters in adaptive networks of phase oscillators,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 29, no. 10, p. 103134, 2019, doi: 10.1063/1.5097835.
- [55] J. Bradbury *et al.*, “JAX: composable transformations of Python+NumPy programs.” [Online]. Available: <http://github.com/jax-ml/jax>
- [56] P. Kidger, “On Neural Differential Equations,” Doctoral dissertation, 2021.
- [57] C. Tsitouras, “Runge–Kutta pairs of order 5 (4) satisfying only the first column simplifying assumption,” *Computers & Mathematics with Applications*, vol. 62, no. 2, pp. 770–775, 2011.
- [58] T. Böhle, C. Kuehn, and M. Thalhammer, “On the reliable and efficient numerical integration of the Kuramoto model and related dynamical systems on graphs,” *International Journal of Computer Mathematics*, vol. 99, no. 1, pp. 31–57, 2022, doi: 10.1080/00207160.2021.1952997.
- [59] P. Kidger, “On Neural Differential Equations.” [Online]. Available: <https://arxiv.org/abs/2202.02435>
- [60] A. A. Sophiya, A. K. Nair, S. Maleki, and S. K. Krishnababu, “A comprehensive analysis of PINNs: Variants, Applications, and Challenges.” [Online]. Available: <https://arxiv.org/abs/2505.22761>
- [61] Y. Bengio, N. Léonard, and A. Courville, “Estimating or Propagating Gradients Through Stochastic Neurons for Conditional Computation.” [Online]. Available: <https://arxiv.org/abs/1308.3432>
- [62] F. Mentzer, D. Minnen, E. Agustsson, and M. Tschannen, “Finite Scalar Quantization: VQ-VAE Made Simple.” [Online]. Available: <https://arxiv.org/abs/2309.15505>
- [63] D. Hafner, J. Pasukonis, J. Ba, and T. Lillicrap, “Mastering Diverse Domains through World Models.” [Online]. Available: <https://arxiv.org/abs/2301.04104>
- [64] Y. Bengio, A. C. Courville, and P. Vincent, “Representation Learning: A Review and New Perspectives,” *CoRR*, 2012, [Online]. Available: <http://arxiv.org/abs/1206.5538>
- [65] J. D. Carroll and P. E. Green, “CHAPTER 3 - Vector and Matrix Concepts from a Geometric Viewpoint,” *Mathematical Tools for Applied Multivariate Analysis*. Academic Press, San Diego, pp. 77–126, 1997. doi: <https://doi.org/10.1016/B978-012160954-2/50004-4>.

- [66] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: MIT Press, 2012.
- [67] M. B. A. McDermott, H. Zhang, L. H. Hansen, G. Angelotti, and J. Gallifant, “A Closer Look at AUROC and AUPRC under Class Imbalance.” [Online]. Available: <https://arxiv.org/abs/2401.06091>
- [68] P. Rockenschaub *et al.*, “Generalisability of AI-based scoring systems in the ICU: a systematic review and meta-analysis,” *medRxiv*, 2023, doi: 10.1101/2023.10.11.23296733.
- [69] P. Kidger and C. Garcia, “Equinox: neural networks in JAX via callable PyTrees and filtered transformations,” *Differentiable Programming workshop at Neural Information Processing Systems 2021*, 2021.
- [70] G. E. H. Jimmy Lei Ba Jamie Ryan Kriso, “Layer Normalization,” *arXiv:1607.06450*, 2016.
- [71] D. Hendrycks and K. Gimpel, “Gaussian Error Linear Units (GELUs).” [Online]. Available: <https://arxiv.org/abs/1606.08415>
- [72] A. Kendall, Y. Gal, and R. Cipolla, “Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics,” *CoRR*, 2017, [Online]. Available: <http://arxiv.org/abs/1705.07115>
- [73] I. Loshchilov and F. Hutter, “Decoupled Weight Decay Regularization.” [Online]. Available: <https://arxiv.org/abs/1711.05101>
- [74] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, “On Calibration of Modern Neural Networks.” [Online]. Available: <https://arxiv.org/abs/1706.04599>