TUHH | UKE

Master Thesis

Juri Backes

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

February 17, 2026

supervised by:
Prof. Dr. Tobias Knopp
M. Sc. Artyom Tsanda

Hamburg University of Technology
Institute for Biomedical Imaging
Am Schwarzenberg-Campus 3
21073 Hamburg

University Medical Center Hamburg-Eppendorf
Section for Biomedical Imaging
Lottestraße 55
22529 Hamburg

Eigenständigkeitserklärung

Hiermit erkläre ich, Juri Backes, an Eides statt, dass ich die vorliegende Masterarbeit im Studiengang "Informatik-Ingenieurwesen" 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 generative Künstliche Intelligenz (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.

Berlin, den 17.02.2026

(Juri Backes)

Acknowledgements

I would like to thank Prof. Dr. Renz and Prof. Dr. Schöll for introducing me to this fascinating research topic, for the numerous discussions, and for their support throughout the creation of this thesis. Furthermore, I would like to thank my supervisors Prof. Dr. Knopp and M. Sc. Tsanda for allowing me to work on this project, for the substantial discussions and inputs, and for creating the opportunity for me to attend an expert meeting on the topic. Lastly, I would like to thank all my friends and family for their support throughout this journey.

Contents

List of tables

List of figures

Acronyms Index

Summary

As the most extreme course of an infectious disease, sepsis poses a serious health threat, with a high mortality rate and frequent long-term consequences for survivors. Despite its enormous burden on global healthcare and ongoing research efforts, early sepsis prediction remains challenging due to the complex nature of its pathophysiology. Current approaches face a fundamental trade-off: data-driven machine learning models achieve strong performance but lack interpretability, while biologically inspired models provide mechanistic insights but have limited clinical validation. This thesis develops the *Latent Dynamics Model*, a hybrid machine learning approach that integrates a functional model of coupled oscillators representing organ- and immune-cell populations. By projecting high-dimensional patient data into the low-dimensional parameter space of the functional model, machine-learned trajectories through this space enable detection and prediction of critical organ system states. The proposed method is trained and evaluated retrospectively on real intensive care patients, achieving state of the art performance. Qualitative analysis reveals that learned trajectories exhibit clinically plausible patterns of deterioration, recovery, and stability. This work demonstrates that embedding biologically grounded structure can improve both predictive performance and interpretability in sepsis prediction.

x

1
Introduction

Nearly 20% of all deaths worldwide and approximately 11 million deaths annually stem from sepsis, a dysregulated response to infection [1]. Despite its enormous clinical and economic burden, sepsis remains notoriously difficult to diagnose in time or even predict in advance due to its inherent complexity. Sepsis is not a single disease but a heterogeneous syndrome involving infection, immune dysregulation, and multi-organ dysfunction across multiple biological scales. Currently, automated early detection systems suffer from limited clinical adoption and further research for improving performance is necessary [2]. Yet early recognition is critical, each hour of delayed treatment increases mortality risk [3].

Given this time-critical challenge, existing computational approaches face a fundamental trade-off: Data-driven methods that use machine learning on electronic health records show promising performance but function as black boxes, offering little mechanistic insight into why a patient is at risk [4, 5]; Mechanistic models explicitly encode biological processes, providing interpretability but require detailed parameterization and high-resolution measurements rarely available in clinical practice [6, 7].

This work proposes a hybrid approach that combines the strengths of both paradigms. The foundation is the Dynamic Network Model (DNM) [8, 9], a functional model that describes sepsis-related organ dysfunction through coupled oscillator networks. The DNM represents immune and organ systems as interacting layers of phase oscillators, where synchronization patterns correspond to physiological states. Two key parameters, a biological age and the strength of organ-immune linkage govern transitions between healthy synchronized states and pathological desynchronized regimes.

While the DNM demonstrates rich theoretical behavior aligned with clinical understanding, no prior work has tested its validity. This thesis addresses that gap by developing the Latent Dynamics Model (LDM), a novel neural architecture that embeds the DNM within a machine learning pipeline for online sepsis prediction.

The LDM learns to map electronic health record time series to interpretable components aligned with the technical sepsis definition. Rather than treating sepsis as a black-box classification problem, the model represents patient organ states as trajectories through the DNMs parameter space, where desynchronization serves as a continuous proxy for acute organ failure. This design provides both competitive predictive performance, evaluated on real-world data and compared to existing baseline methods, and interpretable intermediate outputs that align with clinical reasoning.

The thesis proceeds as follows, building from clinical foundations to methodological innovation: Chapter 2 provides medical background on sepsis, from cellular immune response and its systemic consequences to the Sepsis-3 clinical definition, and closing by establishing

the clinical need for better prediction systems. Chapter 3 reviews the state of the art in sepsis prediction, contrasting mechanistic and data-driven approaches and identifying the gap that motivates hybrid modeling. Chapter 4 introduces the DNM, starting with Kuramoto oscillators and building up complexity to the complete DNM. Numerical simulations demonstrate how the parameter space captures transitions between healthy and pathological states. Chapter 5 presents the LDM methodology in detail, task formalization, architecture design, training objectives, and the differentiable DNM embedding strategy. Chapter 6 describes the experimental setup (data-basis, cohort definition, implementation), presents quantitative results and baseline performance comparisons alongside qualitative analysis of learned patient trajectories. Chapter 7 summarizes findings, discusses limitations, and proposes directions for future research.

2

Medical Background

In 2017, an estimated 48.9 million people worldwide suffered from sepsis and the same year, 11 million deaths were associated with sepsis [1]. This 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* [10]. 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 a chronic disease [11]. But the underlying biological triggers for sepsis as well as individual patient progressions are highly diverse and heterogeneous. Moreover, a septic condition cannot be reduced to a single specific physiological phenomenon, instead it combines multiple complex and interdependent processes across different biological scales. This makes it notoriously difficult to study and diagnose in clinical practice.

Starting with Section 2.1, the pathology and underlying biology are described phenomenologically. Followed by Section 2.2, which provides a more medical and technical overview, this subsection 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" [12]. 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 today.

2.1 Biology of Sepsis

This subsection 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, how this can escalate to *cytokine storms* in Section 2.1.2 and ending with systemic organ failure in Section 2.1.3.

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 [13].

Everything that is 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 tissue's structure, but it also actively participates in biochemical signaling and immune regulation. In this way, it helps to maintain a healthy and balanced tissue, the *homeostasis*, and enables coordinated responses to injury or infection [14].
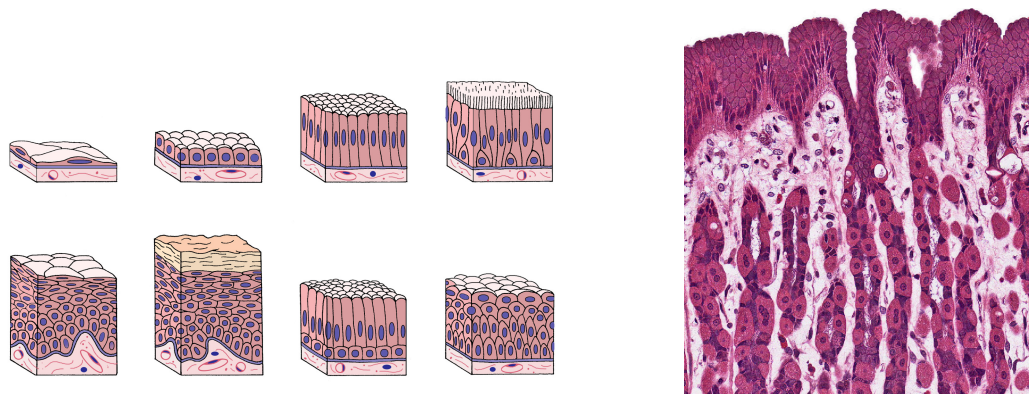
Figure 2.1: On the left-hand side, an illustration of parenchymal cells sitting on top of the basal lamina (thin light-blue section) and the stroma in very light pinkish color building the base. The illustration shows the parenchymal cells in different shapes and organization (image taken from [15]). On the right-hand side a microscopic image of stomach tissue (400x magnification), 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 (image taken from [16]).

Any organism that can cause a disease is called a pathogen, this includes bacteria, fungi, algae, and 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, gets activated [17].

This rapid response does not require the body to have seen the specific pathogen before, as 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) [18]. The PAMPs and DAMPs can be detected by pattern recognition receptors, 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* [19]. 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 act as molecular messengers that coordinate the recruitment of circulating immune cells and will guide them to the location of infection or injury [19]. Additionally, they play a 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.

One specialty of these relatively simple proteins is that they can be produced by almost every other cell, with different cells can produce the same cytokine. Further, cytokines are redundant, meaning targeted cells can show identical responses to different cytokines [20], these features seem to fulfill some safety mechanism to guarantee vital communication flow. After release, cytokines have a relatively short half-life (only a few minutes) but through cascading effects the cytokines can have substantial impact on their micro-environment.

Under normal circumstances, the release of inflammatory cytokines is tightly regulated in time and magnitude. After a pathogen is detected, the release of cytokines is quickly initiated. The release peaks as immune cells are recruited and then automatically fades out once the initial pathogen is controlled and the host returns to a healthy and balanced state, the homeostasis.

2.1.2 Cytokine Storms

In certain scenarios a disturbance to the regulatory mechanisms triggers a positive inflammatory feedback loop, accompanied by a massive release of pro-inflammatory cytokines. These cytokines further activate additional immune and non-immune cells, which in turn amplify the cytokine production, creating a self-reinforcing cycle of immune activation [21]. Ultimately, this leads to a continuous and uncontrolled release of cytokines that fails to shut down, this corresponds to the "host's dysregulated response to an infection" of the Sepsis-3 definition. 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 it almost impossible to distinguish between an appropriate reaction and a harmful overreaction [21]. Out of all cytokines, clinicians can measure only a small subset or secondary markers through blood samples to detect increased cytokine activity. This limited accessibility makes cytokines difficult to study in general, and generally, they prove to be little useful as direct indicators of pathogenesis or for diagnostic purposes.

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 are not fully understood. 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 [22]. Multiple therapeutic inter-

ventions have been tested in clinical trials, yet none have achieved a significant improvement in survival outcomes [18].

2.1.3 Systemic Consequences and Organ Failure

As more cytokines accumulate, they flood not only infected areas, but also surrounding parts of the tissue and circulation, causing the 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* [23], 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 cellular 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 [18].

Step by step, the death of cells spreads throughout the body and affects organ functionality. When multiple organs fail simultaneously, the condition becomes irreversible [12]. 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 pinpoint the exact moment at which the immune response switches from normal to dysregulated behavior. To classify patients as 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 was 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" [12], hence the name Sepsis-3, provides to this day 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" [12], and septic patient progressions 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 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 followed by the cultivation of body fluids, or the other way around, with a SI [12].

   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 antibiotics have to be prescribed in the following 72 hours [12]. 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 SI-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 [12, 24]. Nowadays, the score is 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 assesses 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 organ system.

   Included organ systems are 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 A.1. in the Appendix A.1. The magnitude of a patient's initial SOFA-score captures preexisting organ dysfunction. An increase in SOFA-score ≥ 2, in consecutive assessments, corresponds to an acute worsening of organ functionalities and a drastic worsening in the patients condition, which is used as the indicator to a dysregulated host 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 in SOFA ≥ 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 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 24 hours previous to the increase.

Differences in definitions greatly influence the detection of sepsis, which are used for prevalence estimates for example [25]. Using the lowest SOFA score as baseline, the increase $\geq$ 2 for patients with SI was associated with an 18% higher mortality rate according to [24], a retrospective analysis based on Intensive Care Unit (ICU) data.
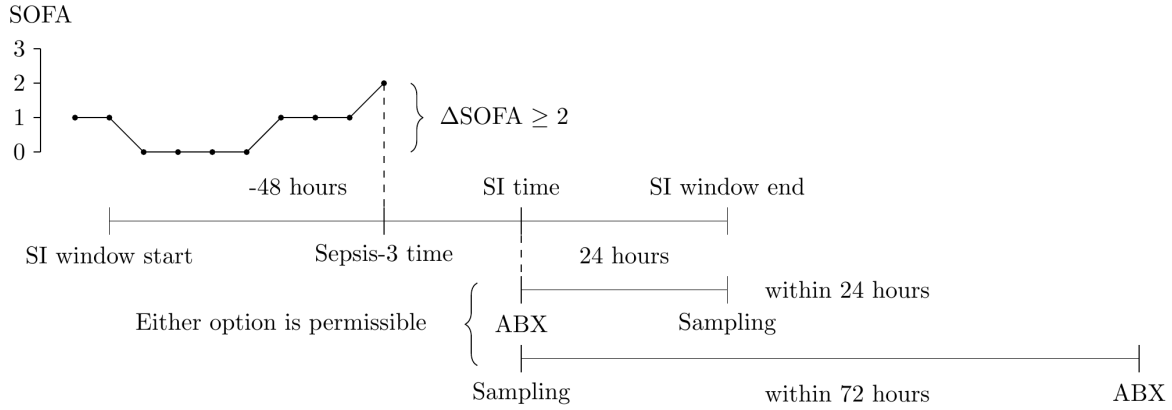
Figure 2.2: Graphical representation of the timings in the Sepsis-3 definition, taken from [26].

2.3 Sepsis Prediction

To this day, sepsis remains as an extreme burden on 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 [27].

Globally, sepsis remains significantly under-diagnosed due to its nonspecific symptoms. Environmental and socioeconomic factors, such as insufficient sanitation, limited access to clean water and healthcare increases the incidence particularly in low- to middle income countries [1, 27].

Traditional sepsis screening has relied on clinical scoring systems such as SOFA or Quick Sequential Organ Failure Assessment (qSOFA), a simplified of the SOFA-score for faster bedside assessment but less validity for in-hospital mortality [24]. 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 organ failure develops. Usually, these alerts are based on predetermined criteria that 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 early warning systems implemented in clinical practice, automating patient screening based on Electronic Health Records (EHRs), found no evidence for improvement in patient outcomes, suggesting insufficient predictive power of analyzed alert systems or inadequate system integration [2]. Nevertheless, positive treatment outcomes depend heavily on timely recognition and intervention [27]. Each hour of delayed treatment increases mortality risk, underscoring the critical importance of early detection [3]. Furthermore, structured screening and early warning systems have demonstrated reductions in time-to-antibiotics and improvements in outcomes [28]. These findings confirm that earlier identification of sepsis improves clinical results and emphasize the need for developing reliable alert systems.

A recent study suggests a paradigm shift in sepsis detection, namely from a purely measurement- and symptom-based to a systems-based approach [29]. 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 [30]. 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.

Yet, the translation of predictive models into clinical practice has proven to be challenging. Implementation 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. Additionally, 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 [30].

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 Chapter 2 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 to 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 [31].

Most existing works focus on dynamic immune response on a cellular level [6, 7, 32, 33]. 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 estimation of the likelihood of sepsis development [32]. More advanced digital twins which incorporate bidirectional feedback between the mechanistic model and patient data have also been explored in [7].

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 the insights 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 [4, 5, 34, 35, 36, 37]. The reviews include a total of 180 studies (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 [4, 5, 34] 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. Both schemes are shown in Figure 3.1, 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, the choice of what time range around a diagnosed sepsis onset qualifies as positive label influences prediction accuracy.
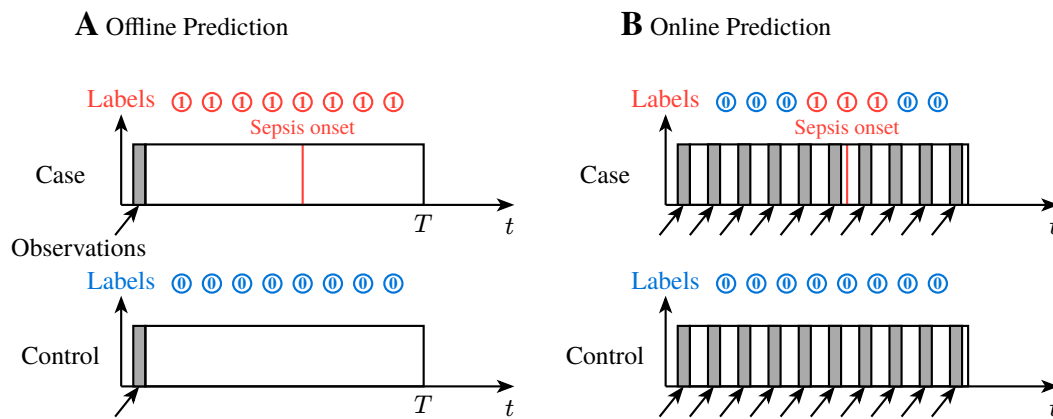
Figure 3.1: Illustration of the two predictions schemes, *offline* (**A**) vs. *online* (**B**) (figure heavily inspired by [5]). The main difference is the sepsis labeling, as well as provision and utilization and arrival of observation data.

Most models rely on routinely collected clinical data, including vital signs, laboratory measurements, demographics, and treatment variables aggregated and summarized in a EHR. Publicly available ICU datasets, for example the Medical Information Mart for Inten-

sive Care (MIMIC) series [38], serve as the predominant development and benchmarking platforms. Differences in feature selection substantially influence 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 ranging in size from 2 to 100, again emphasizing the heterogeneous nature of the field of research.

Most sepsis prediction models are trained retrospectively and evaluated using offline prediction tasks, typically predicting sepsis onset $T = 6$– 48 hours in advance. Model performance is commonly reported using Area Under Receiver Operating 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. For comparison, classical assessment scores achieve AUROCs of SOFA 0.667 and qSOFA 0.612 [34].

Methodologically, a wide range of supervised learning approaches has been explored. Classical models such as logistic regression, Cox proportional hazards models, and random forest 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 relies on Shapley-values analyses, deriving post-hoc importance factors for single input-features or input-feature interactions [37, 39].

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 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 [25]. 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.

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. A major challenge remains in the incomparability and lack of standardization in model development and evaluation. Though, works such as Yet Another ICU Benchmark (YAIB) [40] attempt to address the current 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.

3.3 Summary

Purely model-based and purely data-driven approaches 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, therefore, 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 potentially 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.

After introducing the methodology of this work in Chapter 5, its performance validated on real clinical data in Chapter 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

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 [41].

In essence, cell-to-cell and cell-to-organ interactions in septic conditions form a highly dynamic, nonlinear and spatio-temporal network of relationships [31], which cannot be fully understood by a reduction to single time-point analyses. 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 [42]. 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 [42]. 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 studied the cardiovascular system [43], while more recent studies have focused on the cardio-respiratory coupling [44] and large-scale brain network dynamics [45]. Network approaches have also provided mechanistic insights into disease dynamics, for example Parkinson [46] and Epilepsy [47].

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 [8] and Berner et. al [9] have introduced a dynamical system that aims to model 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: Section 4.1 introduces the theoretical backbone of the 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 systems of coupled oscillators is the *Kuramoto Phase Oscillator Model*. This system of coupled Ordinary Differential Equations (ODEs), 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 $\varphi_i \in \mathbb{R}$, for $i = 1, ..., N$ and an intrinsic frequency $\omega_i \in \mathbb{R}$ [49]. The phase φ_i represents the angular position of oscillator i and evolves as a real number. The sine coupling ensures the dynamics are 2π-periodic, allowing mapping to the unit circle for visualization. Traditionally, two oscillators i and j are coupled through the sine function $\sin(\varphi_i - \varphi_j)$, the dynamics are then given by the ODE system:

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

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 $\mathrm{K} \in \mathbb{R}_{>0}$ between oscillators i and $j \in 1...N$. 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 the system evolves 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 4.1. 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].
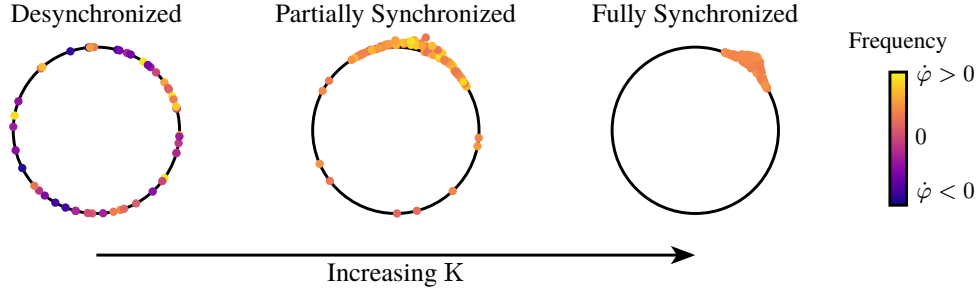
Figure 4.1: Illustration of the collective dynamics of a population of phase oscillators on the unit circle as the global coupling strength K is increased. Each point represents an oscillator at phase φ_i, color encodes the instantaneous phase velocity $\dot\varphi_i$ (blue/yellow: faster, green: slower). 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 both 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 α, 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:

$$\dot\varphi_i = \omega_i - \frac{K}{N} \sum_{j=1}^{N} \sin(\varphi_i - \varphi_j + \alpha) \tag{4.2}$$

As a result the critical coupling strength K_c increases with α [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 $\mathbf{K} = (\kappa_{ij})_{(i,j) \in N \times N}$:

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

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

Here, the adaptation 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 phase lag parameter β 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 κ_{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, at 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 $\mu = 1, ..., L$ and $\nu = 1, ..., L \neq \mu$) via interlayer links:

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

Here μ and ν 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 α and β and sample distribution of initial phases.

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 of multiplexed layers changes the system behavior once more, for example individual 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 states, for example, one in a clustered the other in a splay state.

4.2 Description of the Dynamic Network Model

Built on top of the Kuramoto Oscillator Model, the DNM is a **functional** model, that means it **does not try to model accurately on any cellular, biochemical, or organ level**, it instead tries to model dynamic interactions. At the core, the model differentiates 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 patient's 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 $\cdot^1$ superscript and is called the *organ layer*, stroma cells are indicated with $\cdot^2$ and is referred to as nonspecific *immune layer*. Each layer consists of N phase oscillators $\varphi_i^{1,2} \in \mathbb{R}$. 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 4.2.
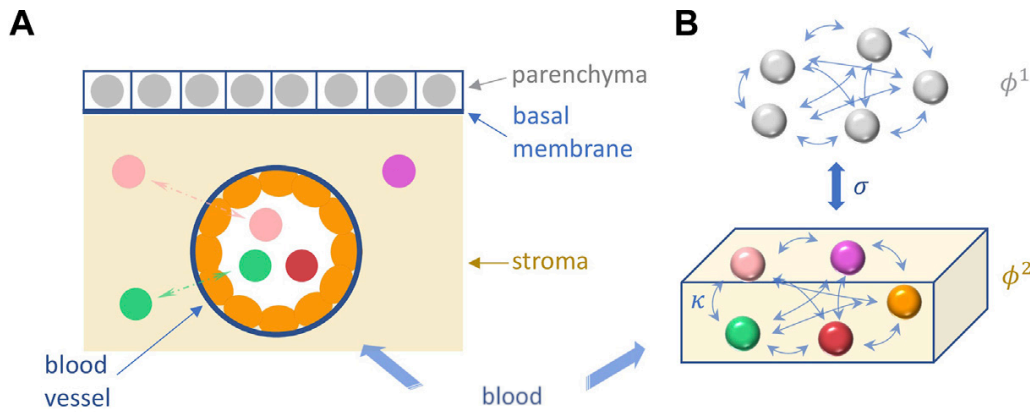
Figure 4.2: Image taken from [9] in a simplified form, illustrating the Dynamic Network Model. **A** A tissue element is depicted, in which the basic processes of sepsis take place. Shown are different immune and structural cells involved (colored) in the stroma (yellow area), the parenchyma (grey), 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 $\boldsymbol{A}^1 \in [0,1]^{N \times N}$ with elements a_{ij}^1, representing a fixed connectivity within an organ.

The dimensionless system dynamics are described with the following coupled ODE terms, built 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 flow 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 α^{11} and α^{22} modeling cellular reaction delay. To separate the fast moving oscillator dynamics from the slower moving coupling weights adaptation rates $0 < \varepsilon \ll 1$ are introduced. Since the adaptation of parenchymal cytokine communication is assumed to be slower than the immune counterpart [8], it is chosen $\varepsilon^1 \ll \varepsilon^2 \ll 1$, which introduces dynamics on multiple timescales.

Lastly, the most influential parameter is β which controls the adaptivity of the cytokines. Because β 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 [9].

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
Variables		
φ_i	Phase	Group of cells
$\dot{\varphi}_i$	Phase Velocity	Metabolic activity
κ_{ij}	Coupling Weight	Cytokine activity
Parameters		
α	Phase lag	Metabolic interaction delay
β	Plasticity rule	Combination of risk factors
ω	Natural frequency	Natural cellular metabolism
ε	Time scale ratios	Temporal scale of cytokine activity
a_{ij}	Connectivity	Fixed intra-organ cell-to-cell interaction
σ	Interlayer coupling	Interaction between parenchymal and immune cells through the basal lamina
Measures		
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 [14]. 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 healthy equilibrium state 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 this synchronicity and the formation of frequency clusters in the parenchymal layer, i.e. loss of homeostatic balance. In the DNM at 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 [9]. 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 patient's 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 [9], since they will be serving as a basis for following chapters.

4.3.1 Numerical Integration

The backbone of 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. These feature allow high-level numerical code to be translated to highly optimized accelerator-specific machine code, including graphics cards. 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 [9] uses a fourth-order Runge-Kutta method and a fixed step-size, this implementation[1], in contrast, uses the Tsitouras 5/4 Runge-Kutta method [57] with adaptive step-sizing controlled by a proportional-integral-derivative 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 of the form $\sin(\theta_l - \theta_m)$ were expanded as:

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

---

[1]The code is available at https://github.com/unartig/sepsis_osc/blob/main/src/sepsis_osc/dnm/dynamic_network_model.py

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.

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 [8] and [9] 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 ω just changes the frame of reference (co-rotating frame), the dynamics stay unchanged [9]. 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 [9] yielding the desired dynamical properties. The constant intralayer coupling in the parenchymal layer 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 [9] the influence of parameter values for β and σ 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 β was increased to $[0.0, 1.0\pi]$. A 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
Variables			
φ_i^1	$\sim \mathcal{U}[0, 2\pi)$	$\kappa_{i \neq j}^1$	$\sim \mathcal{U}(-1, 1)$
φ_i^2	$\sim \mathcal{U}[0, 2\pi)$	$\kappa_{i \neq j}^2$	clusters of size C and $1 - C$
Parameters			
M	50	N	200
C	20%		
β	$[0.0, 1.0]\pi$	σ	$[0.0, 1.5]$
α^{11}, α^{22}	-0.28π	α^{12}, α^{21}	0.0
ω_1, ω_2	0.0	A^1	$\mathbb{1} - I$
ε^1	0.03	ε^2	0.3

Initial values for the system variables, i.e. the phases and coupling strengths, were not parametrized 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 of the intralayer coupling matrix is set to 0 $\kappa_{ij}^{1,2} = 0$.

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 [9], the cluster size $C \in [0, 0.5]$ was chosen as 0.2, but as their findings suggest the size of the initial 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 4.3.
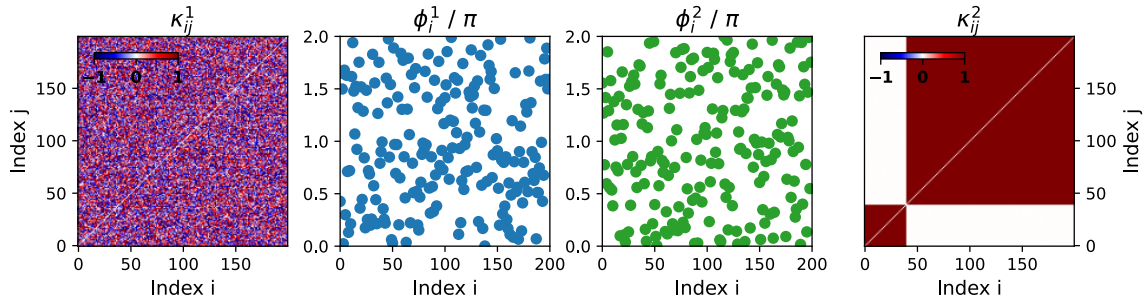
Figure 4.3: Initializations for the variable values of a DNM with $N = 200$ oscillators per layer. The middle two plots show the phases of the oscillators, with φ_i^1 for parenchymal and φ_i^2 for the immune layer, sampled from a uniform random distribution from 0 to 2π. On the left-hand side is the 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.

4.3.3 Synchronicity Metrics

As introduced in Section 4.1, for the 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]:

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

where $R_2^{1,2} = 0$ corresponds to total desynchronization, the splay-state and $R_2^{1,2} = 1$ 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:

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

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

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

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

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. While the overall systematics stay the same, deviations from their original results are expected, due to this change in definition.

One can now calculate the *standard deviation of the phase velocity*:

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

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 $m = 1, ..., M$ with the same parameterization, it is expected that different initializations, meaning initial-values drawn from the parameterized

distributions, exhibit dissimilar behaviors, one can also calculate the *ensemble averaged standard deviation of the phase velocity*:

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

In [9] 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 [9] identify β, the combined age parameter, and σ, the interlayer coupling strength which models the communication between organ and immune cells, 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 parameterizations and initializations. Afterwards, the transient and temporal behavior of the metrics $s^{1,2}$ and $R^{1,2}$ is presented for the same parameterization, as well as the introduction of the β, σ phase space of these metrics.

In Figure 4.4, snapshots of the system variables are shown for different parameterization, differing only in the choice β and σ, 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 of an individual patient.

All the 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: Exemplary β-σ combinations to illustrate simulation results.

	A	**B**	**C**	**D**
β	0.5π	0.58π	0.7π	0.5π
σ	1.0	1.0	1.0	0.2

In Figure 4.4, 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 $\mathbf{K}^2$. Individual oscillators of each layer are sorted first from lowest to highest frequency and secondary by lowest to highest phase for better clarity. Figure 4.4/C and Figure 4.4/C' share the same parameterization but are different samples from the same initialization distributions.
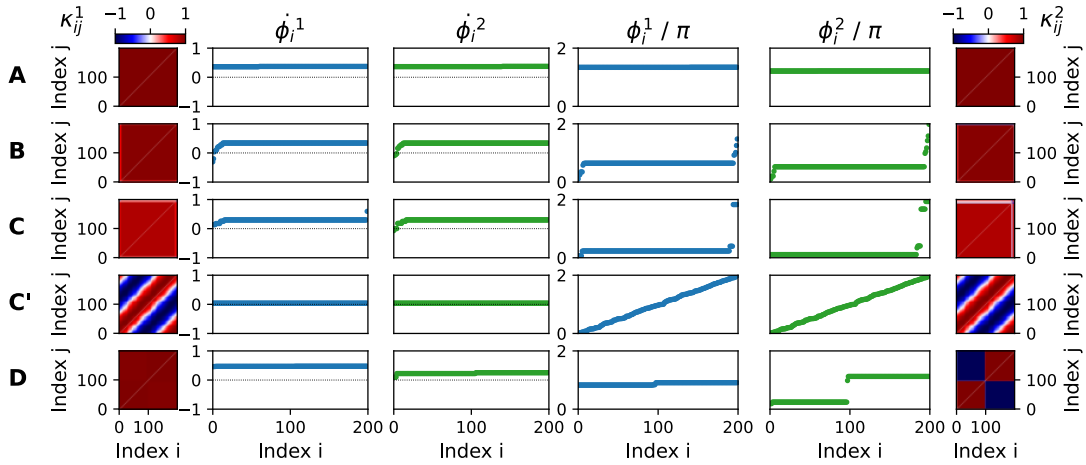
Figure 4.4: Snapshots of different DNM parametrization and initialization. Configuration **A** can be regarded as healthy, with phases and frequencies being fully synchronized. In contrast, **B** and **C** are pathologic, due to their clustering in $\dot{\varphi}^1$. Configuration **C'** corresponds 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 $\dot{\varphi}^1$ are synchronized.

Figure 4.4/A 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. Figure 4.4/B and Figure 4.4/C in contrast show signs of a pathological state, here both the frequencies and phases have distinct clusters. The clusters are also slightly visible in the coupling matrices, where the coupling strength differs based on the cluster, with the smaller clustering having exhibiting a weaker coupling. Figure 4.4/C', even though having the same parameterization as Figure 4.4/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 Figure 4.4/C' show traveling waves, which are characteristic for splay states. Observing different results for different initializations justifies the introduction of ensembles. Lastly row Figure 4.4/D shows a resilient state, where the phases are clustered while the frequencies are synchronized.

In Figure 4.5, where 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 columns show the standard deviation of the mean phase velocities σ_χ for each ensemble member $m \in 1, 2...M$ (blue) and the aggregated ensemble mean $s^{1,2}$ (orange). The plots show the temporal evolution of metrics for quantifying phase and frequency coherence, with the two right-most columns of Figure 4.5 show the temporal behavior of the Kuramoto Order Parameter for each individual ensemble member m (blue) and the aggregated ensemble mean (orange). 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.
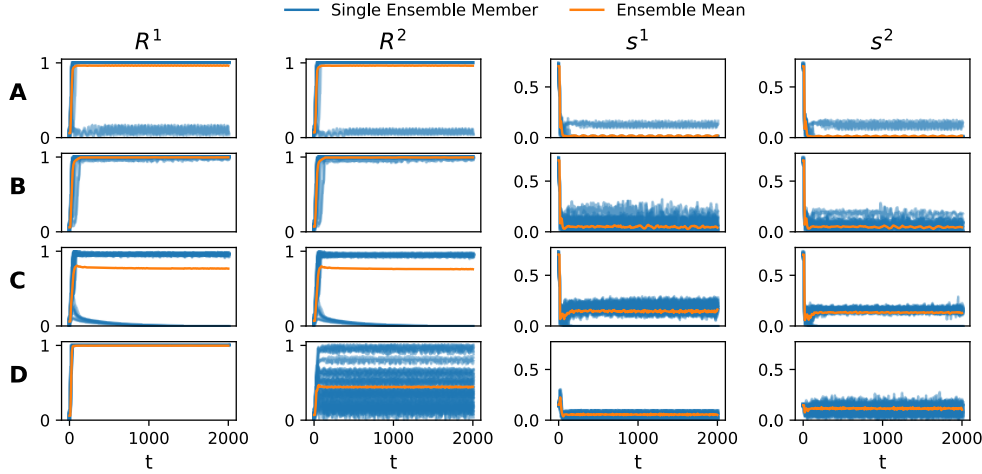
Figure 4.5: Transient and temporal evolution of the phase- and frequency-synchronization metrics $R^{1,2}$ and $s^{1,2}$, for ensembles of the DNM for the configurations listed in Table 4.3. Emphasizing the influence of β and σ on the systems synchronization behavior, and presenting different stable emergent system states. Single ensemble members are shown in blue and the ensemble mean in orange.

Every ensemble in Figure 4.5 shows decoherence for early time-points, which is expected for randomly initialized variables, but changes quickly through a transient phase $t \in [0.0, 200]$ into systematic stable behavior. In Figure 4.5/A, the stable states align with the observations made for Figure 4.4/A, besides small jitter the majority of ensemble members exhibits synchronized frequencies $s^{1,2} \approx 0$. Also the phases of configuration Figure 4.5/A are mostly synchronized with the ensemble mean $\overline{R}^{1,2} \approx 1$, i.e. $\overline{R}^{1,2} = \frac{1}{M} \sum_m^M R_m^{1,2}$ with $M = 50$, only two initializations (ensemble members) 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 Figure 4.5/B, the amount of incoherence inside the ensemble is clearly visible, with mean $s^{1,2}$ being positive and some more ensemble members exhibiting clustering, indicated by a Kuramoto Order Parameter slightly less than 1. In configuration Figure 4.5/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 Figure 4.5/D the overall phase incoherence is again a bit less compared to Figure 4.5/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.

Each of the configurations only differs in the parameter choices for β and σ, yet they evolve into unique and distinct system states. To put these four specific configurations into broader perspective, a grid of β and σ 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 4.6 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 $\overline{R}^{1,2}$ where darker colors indicate larger phase desynchronization. The white rectangle indicates the simulated region

in [9], $\beta \in [0.4, 0.7]$ and $\sigma \in [0, 1.5]$ for reference. Coordinates of configurations A, B, C, and D are also labeled.
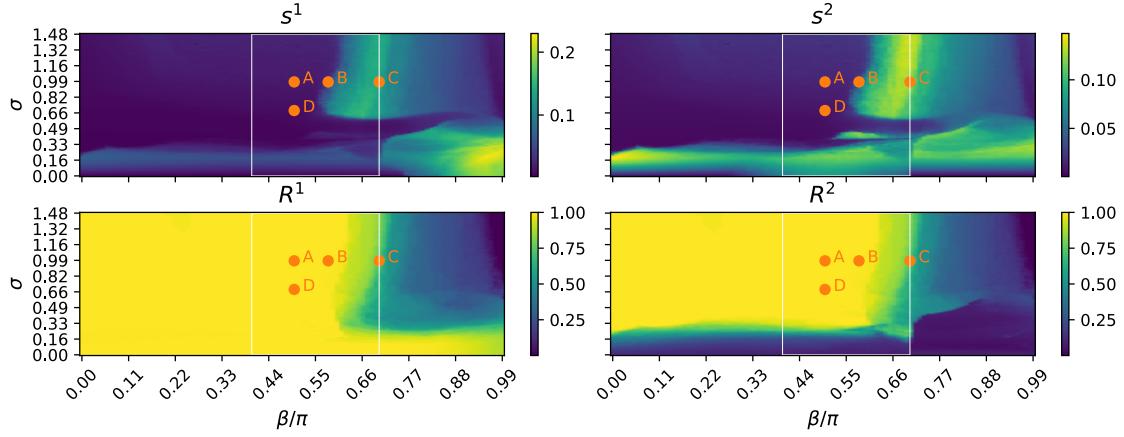
Figure 4.6: The phase space of the parameters β and σ and illustrate the broader picture their influence on the frequency and phase synchronization of the DNM. White rectangle indicates the grid-limits of the original publication [9] .

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 β 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 β.

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 [9]. 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 β the medical age, σ the interlayer coupling, the phase lags α, and the timescale ratios ε 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 (β, σ) 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 β and σ 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

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 and 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 β and interlayer coupling σ were shown to modulate these regimes. However, the gap between theoretical considerations and clinical validation remains unaddressed. 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 data-driven and model free approaches.

5

Latent Dynamics Model

This chapter introduces the LDM, a novel methodological framework that embeds the DNM (Chapter 4) within a machine learning pipeline. As established in the previous chapter, the DNM provides a theoretical framework for understanding sepsis dynamics through coupled oscillator networks, though it has not yet been validated against real patient data. To address this gap, the LDM was designed to infer interpretable and patient specific DNM parameters from clinical observations and simultaneously generate short-term sepsis risk prediction in an online fashion.

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. Namely, the SI and increase in SOFA scores are predicted as proxies creating more nuanced and more interpretable prediction results.

At its core, the LDM embeds the DNM into a learnable latent dynamical system, to provide an inductive bias in the prediction path of the SOFA-score increase. A neural network learns to position patients within the (β, σ) parameter phase space of the DNM and a RNN learns to predict the drift through that space based on observed clinical time series. Ultimately, the goal is to align the predicted trajectories in the parameter space with real patient physiological progressions.

This chapter proceeds with Section 5.1, where the prediction task prediction strategy is 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. The chapter will close with explaining some established performance metrics in Section 5.5.

5.1 Formalizing the Prediction Problem

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.

Even though the presented work is a proof-of-concept study, the prediction system is designed for real-world clinical use, where causality has to be obeyed. This means, 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 $t \in \{1, 2, ..., T\}$ denote an observation time during a patient's ICU-stay of length T and the available EHR at that time consisting of $D \in \mathbb{N}_{>0}$ variables. After imputation of missing values, normalization, and encoding of non-numerical quantities, each variable μ_j is mapped to a numerical value:

$$\mu_{t,j} \in \mathbb{R}, \quad j = 1, ..., D \tag{5.1}$$

These values are collected into a column-vector:

$$\boldsymbol{\mu}_t = \left(\mu_{t,1}, ..., \mu_{t,D}\right)^T \in \mathbb{R}^D \tag{5.2}$$

where the superscript $\cdot^T$ denoting a transpose operation. The vector $\boldsymbol{\mu}_t$ is the most complete description of the ICU-patient's physiological state at observation time point t given the available data. 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 to 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. 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 patient's ICU stay as:

$$S_t := A_t \wedge I_t \tag{5.3}$$

Here $A_t = \{\Delta O_t \geq 2\}$ indicates an acute worsening in organ function, measured via a change in SOFA-score $\Delta O_t = O_t - O_{\mathrm{base}}$ with respect to a patient specific baseline SOFA-score O_{base}. The choice of O_{base} has to align with the Sepsis-3 definition, for example a 24 hours running minimum, the or $O_{\mathrm{base}} = O_{t-1}$, the acute change between two consecutive SOFA-score assessments. The event I_t is an indicator for a SI at time t defined according to the Sepsis-3 definition, spanning the 48 hours before and 24 hours after 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 $\boldsymbol{\mu}_{1:t}$, 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 $\boldsymbol{\mu}_{1:t}$ the target probability is given by:

$$\Pr(S_t|\boldsymbol{\mu}_{1:t}) = \Pr(A_t \wedge I_t \mid \boldsymbol{\mu}_{1:t}) \tag{5.4}$$

5.1.3 Heuristic Scoring and Risk Estimation

The direct estimation of the true conditional probability $\Pr(S_t|\boldsymbol{\mu}_{1:t})$ is both difficult to derive analytically and computationally expensive 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:

$$\Pr(A_t \cap I_t|\boldsymbol{\mu}_{1:t}) = \Pr(I_t|\boldsymbol{\mu}_{1:t})\Pr(A_t|\boldsymbol{\mu}_{1:t}) \tag{5.5}$$

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 $\overline{I}_t \in [0,1]$ 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:

$$\tilde{A}_t \approx \Pr(A_t|\boldsymbol{\mu}_{1:t}) \quad \text{and} \quad \tilde{I}_t \approx \overline{I}_t \tag{5.6}$$

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

$$\Pr(S_t|\boldsymbol{\mu}_{1:t}) \approx \tilde{S}_t := \tilde{A}_t\tilde{I}_t \tag{5.7}$$

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 but an empirical risk score based on the Sepsis-3 definition, serving as a differentiable surrogate for the Sepsis-3 sepsis onset criterion $P(S_t|\boldsymbol{\mu}_{1:t})$. As a heuristic risk score, $\tilde{S}_t$ is a monotonic indicator, where larger values correspond to higher expected risk of sepsis outbreak.

5.1.4 From EHR to Risk Scores

The high-dimensional EHR history $\boldsymbol{\mu}_{1:t}$ 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:

$$\tilde{I}_t = f\big(\boldsymbol{\mu}_{1:t}; \theta_f\big) \tag{5.8}$$

where $f\big(\cdot; \theta_f\big)$ represents a neural network with learnable parameters θ_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:

$$\hat{O}_t := g\big(\boldsymbol{\mu}_{1:t}; \theta_g\big) \tag{5.9}$$

where $\hat{O}_t$ denotes a latent, differentiable estimation of the ground truth SOFA-score O_t. The function $g\big(\cdot; \theta_g\big)$ represents a combined DNM pipeline, where θ_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:

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

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

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

yields a monotonic indicator (larger predicted SOFA-score increase leads to 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 $\Pr(A_t \mid \boldsymbol{\mu}_{1:t}) = 0.5$, making it an arbitrary monotonic transformation rather than a calibrated probability. Yet, this formulation is differentiable and monotonic, therefore enables gradient-based optimization and through end-to-end training on actual sepsis outcomes, a model can to make $\tilde{A}_t$ a useful risk indicator.

5.2 Architecture

The previous subsection explained sepsis onset target S_t can be decomposed into the conjunction of SI indication I_t and an acute organ deterioration event A_t, which itself can be calculated from two consecutive SOFA-scores. The presented LDM is designed to estimate

the fundamental components $\tilde{O}_t$ and $\tilde{I}_t$ from a history of EHR $\boldsymbol{\mu}_{1:t}$ to derive the heuristic sepsis risk score $\tilde{S}_t \approx S_t$ for individual patients.

In the following subsections, two estimator-functions with learnable parameters will be introduced, generating estimations for target components $\tilde{I}_t$ and $\tilde{A}_t$. Each component is estimated by a module built around RNN, e.g. Gated Recurrent Unit (GRU) or Long Short-Term Memory (LSTM), enabling continuous estimation updates based on newly arriving measurements. By keeping these modules fully differentiable with respect to their inputs allows for optimization via first order gradient descent methods.

This section starts with introduction of the estimator module for the suspected infection indication 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 5.1 provides a list of symbols used throughout this chapter. Unless otherwise specified, patient indices are omitted for notational clarity. For example, $\boldsymbol{\mu}_t$ denotes the EHR vector at time t, which formally corresponds to $\boldsymbol{\mu}_{i,t}$ for patient $i \in 1, ..., N$. Explicit patient indices are retained in contexts where summation over patients is required, such as in the loss functions.

Table 5.1: Summary of notation used in the Latent Dynamics Model.

Symbol	Description
$i \in 1, ..., N$	Patient index
$t \in 1, ..., T_i$	Time point index and trajectory length of patient i
$\boldsymbol{\mu}_t \in \mathbb{R}^D$	EHR vector with D variables at time t
$\boldsymbol{\mu}_{1:t}$	EHR history from time 1 to t
$S_t, A_t, I_t \in \{0,1\}$	Binary sepsis onset, organ failure, and infection events
$O_t, \Delta O_t \in 0, ..., 24$	SOFA-score and change from baseline
$\overline{I}_t \in [0,1]$	Continuous surrogate infection indicator
$\hat{O}_t \in \{0, ...24\}$	Estimated SOFA- score
$o_{s,d}(\hat{O}_{t-1:t}) \in (0,1)$	Increase detection function
$\tilde{S}_t, \tilde{A}_t, \tilde{I}_t \in (0,1)$	Predicted estimates for sepsis, organ failure, and infection risks
$\boldsymbol{z} = (z_\beta, z_\sigma)$	Latent coordinates in DNM parameter space
$\hat{z}_t, \Delta\hat{z}_t$	Predicted latent position and change
$s^1(\boldsymbol{z})$	Desynchronization in DNM
$\boldsymbol{h}_{f,g} \in \mathbb{R}^{H_{f,g}}$	Hidden state vector
$f_{\theta_f}, g^e_{\theta_g}, g^r_{\theta_g}, d_{\theta_d}$	Infection indicator, encoder, recurrent, and decoder modules
θ	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
λ_{loss}	Loss weight for component "loss"
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 $\overline{I}_t$, 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:

$$\left(\boldsymbol{\mu}_t, \overline{I}_t\right), \quad t = 1...T_i, \quad i = 1...N, \tag{5.12}$$

a parameterized nonlinear recurrent function

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

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

$$\left(\tilde{I}_t, \boldsymbol{h}_t^f\right) = f_{\theta_f}\left(\boldsymbol{\mu}_t, \boldsymbol{h}_{t-1}^f\right) \tag{5.14}$$

The hidden state $\boldsymbol{h}_t$ propagates temporal information through time. For the first time-step $t = 1$ a learnable initial hidden state $\boldsymbol{h}_1^f$ is used, shared for all patients.

The model is implemented as a RNN, illustrated in Figure 5.1. At each timestep, a GRU-cell updates the hidden state, and a learned linear projection $\boldsymbol{W}_f \boldsymbol{h}_t^f$, with $\boldsymbol{W}_f \in \mathbb{R}^{1 \times H_f}$, followed by sigmoid activation produces the infection risk estimate:

$$\boldsymbol{h}_t = \text{GRU-Cell}_{\theta_f^{\text{gru}}}\left(\boldsymbol{\mu}_t, \boldsymbol{h}_{t-1}^f\right) \tag{5.15}$$

$$\tilde{I}_t = \text{sigmoid}\left(\boldsymbol{W}_f \boldsymbol{h}_t^f + b_f\right) \tag{5.16}$$

where the bias $b_f \in \mathbb{R}$ is a single scalar and $\theta_f = \left\{\theta_f^{\text{gru}}, \boldsymbol{W}_f, b_f, \boldsymbol{h}_f\right\}$ combines all learnable parameters.

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

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

is minimized and thus the estimator provides a differentiable estimate of the surrogate infection activity.
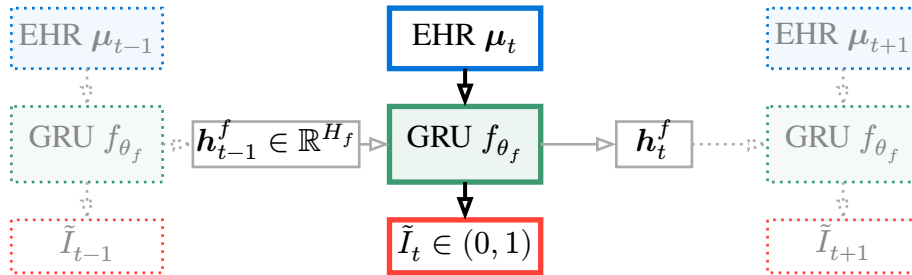
Figure 5.1: Schematic of the Infection Indicator Module architecture and rollout. The RNN process the EHR sequence $\boldsymbol{\mu}_{1:t}$ step-by-step, propagating $\boldsymbol{h}_t$ to capture temporal dependencies, and outputs infection risk estimates $\tilde{I}_t$ at each timestep.

5.2.2 SOFA Predictor Module

The complete SOFA predictor module g_θ is composed two submodules, an initial-encoder g_θ^e and a recurrent latent predictor g_θ^r, each described below. The idea is to translate the physiological patient trajectory to a sequence of the DNM parameters β and σ, 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. Afterwards, 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

Recall that 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 $z = (z_\beta, z_\sigma) = (\beta, \sigma)$, keeping every other system parameter according to Table 4.2, yields the corresponding amount of desynchronization $s^1(z)$. The value $s^1(z)$ corresponds to the amount of desynchronicity at the end of the integration time $s^1(T_{\text{sim}})$ at the coordinates of z. The time index of s_t^1 is omitted for readability. Given a desynchronization measure $s^1(z)$, the SOFA approximation $\hat{O}(z)$ is calculated using:

$$\hat{O}(z) = \left\lfloor \frac{24 \cdot s^1(z)}{s_{\text{max}}^1} \right\rceil = \left\lfloor \frac{24 \cdot s^1(\beta, \sigma)}{s_{\text{max}}^1} \right\rceil \tag{5.18}$$

The space spanned by the two parameters is called the *latent space*, coordinate-pairs in that latent space are denoted $z = (z_\beta, z_\sigma)$. In this work only a predefined subspace of the entire (β, σ) plane is used. s^1 values are normalized to the interval $[0, 1]$ by dividing by the maximum value within the subspace s_{max}^1, which allows retrieval of all 24 SOFA levels. 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 on the latent location and additional clinical information, the patient will perform a trajectory through the latent space yielding step-by-step SOFA-score $\hat{O}_{1:t}(z)$ estimates needed to calculate the heuristic organ failure statistic $\tilde{A}_t$.

5.2.2.2 Latent Parameter Dynamics

An neural encoder connects the initial $t = 1$ and high-dimensional EHR-vector $\boldsymbol{\mu}_1$ to the dynamical regime of the DNM:

$$g^e_{\theta^e_g} : \mathbb{R}^D \to \mathbb{R}^2 \times \mathbb{R}^{H_g} = \mathbb{R}^{2+H_g} \tag{5.19}$$

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

$$\left(\hat{\boldsymbol{z}}^{\mathrm{raw}}_1, \boldsymbol{h}^g_1\right) = \left(\left(\hat{z}^{\mathrm{raw}}_{1,\beta}, \hat{z}^{\mathrm{raw}}_{1,\sigma}\right), \boldsymbol{h}^g_1\right) = g^e_{\theta^e_g}(\boldsymbol{\mu}_1) \tag{5.20}$$

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. In addition to the estimated latent coordinate $\hat{z}^{\mathrm{raw}}_1$, 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 $\boldsymbol{h}^g_1 \in \mathbb{R}^{H_g}$ is the initial hidden space.

To keep latent coordinates in the predefined area they are ultimately transformed by:

$$\hat{\boldsymbol{z}} = \mathrm{sigmoid}(\hat{\boldsymbol{z}}^{\mathrm{raw}}) \cdot \begin{pmatrix} \beta_{\max} - \beta_{\min} \\ \sigma_{\max} - \sigma_{\min} \end{pmatrix}^T + \begin{pmatrix} \beta_{\min} \\ \sigma_{\min} \end{pmatrix}^T \tag{5.21}$$

Where $\cdot$ is the element wise matrix multiplication. The latent coordinate $\hat{z}_1$ 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, see Equation (5.18).

Since the heuristic SOFA risk $\tilde{A}$ depends on the evolution of organ function $\hat{O}_{1:t}$, it is necessary to estimate not only the initial state $\hat{z}_1$ but also its evolution. For this purpose a neural recurrent function:

$$g^r_\theta : \mathbb{R}^{D+2} \times \mathbb{R}^{H_g} \to \mathbb{R}^2 \times \mathbb{R}^{H_g} \tag{5.22}$$

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

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

$$\boldsymbol{h}_t = \mathrm{GRU\text{-}Cell}_{\theta^{\mathrm{gru}}_g}\left((\boldsymbol{\mu}_t, \hat{z}^{\mathrm{raw}}_{t-1}), \boldsymbol{h}^g_{t-1}\right), \quad t = 2, ..., T \tag{5.23}$$

$$\Delta\hat{z}^{\mathrm{raw}}_t = \boldsymbol{W}_g \boldsymbol{h}^g_t \tag{5.24}$$

$$\hat{z}^{\mathrm{raw}}_t = \hat{z}^{\mathrm{raw}}_{t-1} + \Delta\hat{z}^{\mathrm{raw}}_t \tag{5.25}$$

where $\theta^r_g = \{\theta^{\mathrm{gru}}_g, \boldsymbol{W}_g\}$ 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 and should translate to the pathological evolution of patients. The online-prediction rollout is shown in Figure 5.2.

The model learns incremental changes in the latent space Δz_t rather than absolute positions z, this naturally constraints the magnitude of changes and encourages the prediction of more gradual trajectories of the physiological state within the latent space. 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 g_θ^e and g_θ^r, the latent points z are positioned to minimize the Mean Squared Error (MSE) loss:

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

where the class-balancing weight:

$$w_O = \log\left(1 + f_O^{-1}\right) \tag{5.27}$$

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 $[0, 1]$.
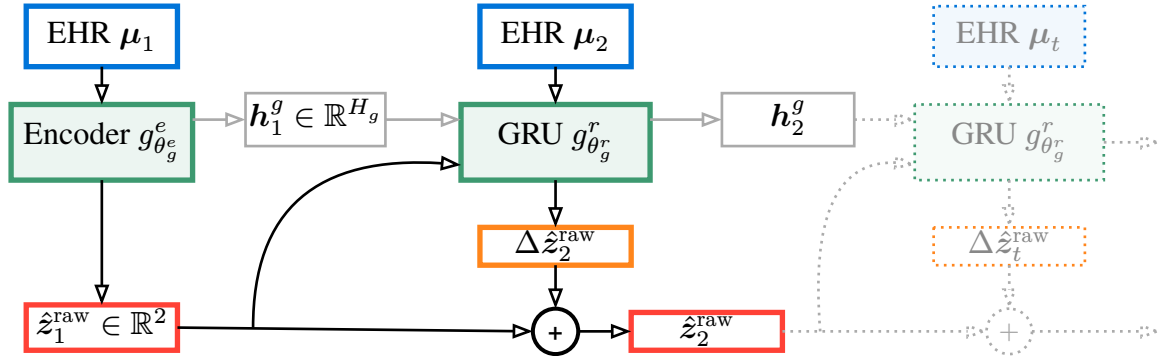
Figure 5.2: Schematic of the online-prediction rollout by the SOFA predictor module. The Encoder $g_{\theta_g^e}$, generates the initial latent position $\hat{z}_1^{\text{raw}}$ based on the first observed EHR. Afterwards, the RNN processes the following EHR sequence $\mu_{1:t}$ step-by-step, maintaining 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 $\hat{z}_{t-1}^{\text{raw}} + \Delta \hat{z}_t^{\text{raw}}$

Theoretically, gradients can flow backwards through the whole sequence, jointly train the encoder g_θ^e and recurrent function g_θ^r by minimizing the loss $\mathcal{L}_{\text{sofa}}$. Computing $s^1(z)$ requires numerically integrating the coupled ODE system of the DNM, which is expensive for each forward pass. More problematically, backpropagating gradients through this integration over a long simulation time T_{sim} and across an ensemble of M system, each consisting of N oscillators, will likely lead to vanishing gradients and numerical instability. This would provide uninformative learning signals to the latent coordinates (β, σ). Motivated by these challenges,

the next subsection introduces a quantization-based lookup strategy which precomputes the DNM dynamics and maintains differentiability through interpolation.

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 (β, σ). This approach is taken in Neural Differential Equations [56] and Physics Informed Neural Networks [59] where gradients are typically backpropagated through the ODE integration to their input parameters (β, σ 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 caching methodology that still provides meaningful gradients and simultaneously reduces the computational burden. For that, the continuous latent space is quantized to a discrete and regular grid, with the metric s^1 precomputed for each coordinate pair in a 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 [9]). To retrieve values, 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 (β, σ)-space, the quantized metrics are interpolated by smoothing nearby quantization points with a Gaussian-like kernel, which is illustrated in Figure 5.3.

To enable gradient-based optimization, i.e. being differentiable, the lookup of nearby points z' combines two mechanisms. Firstly, a straight-through estimator [60], for the discrete grid indexing operation, allowing gradients to flow as if the rounding were identity.

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

In the forward pass, this equals the rounded value for lookup, in the backward 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:

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

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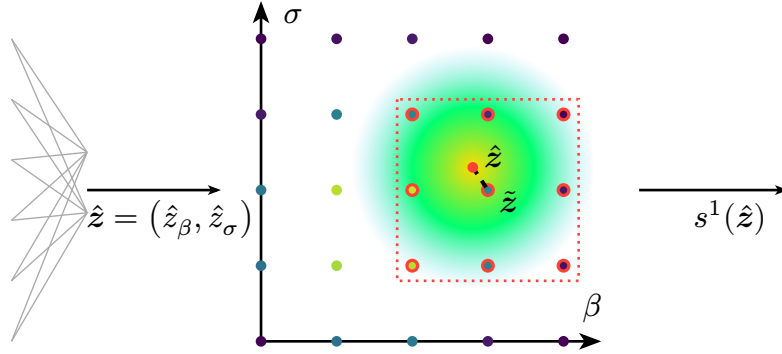
While the squared distances ($\|\hat{z} - z'\|^2$) receive exponentially more weight, the softmax operation normalizes the weights to 1, creating a convex combination of weights:

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

The K neighboring points can be calculated via:

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

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

Figure 5.3: Quantized latent lookup of precomputed synchronization metrics. Point colors represent the amount of desynchronization s^1 in the parenchymal layer. Neighboring points, the $\boldsymbol{z}' \in \mathcal{N}_{3 \times 3}(\tilde{\boldsymbol{z}})$ sub-grid, indicated by the red outlines and the red rectangle around $\tilde{\boldsymbol{z}}$, are used smoothed using a Gaussian-like kernel, represented by the color gradient around estimation point $\hat{\boldsymbol{z}}$. This allows continuous interpolation the parameter space.

This quantization strategy, called *latent lookup*[2] is closely related to *finite scalar quantization* [61], used in Dreamer-V3 [62] for example. In contrast to this presented latent lookup, the latent coordinates in Dreamer-V3 do not have prior semantic meaning associated with them, yet both allow for differentiable quantization. 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 4.6, multiple latent coordinates $\boldsymbol{z}$ 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.

Following the classical Auto-Encoder framework [63], a decoder module is added as an auxiliary regularization component during training to encourage a semantically structured latent space beyond SOFA-scores. This latent regularization is motivated by *Representation Learning* [63], and ensures that nearby points in the latent (β, σ)-space correspond to

---

[2]Implementation is available at https://github.com/unartig/sepsis_osc/blob/main/src/sepsis_osc/ldm/lookup.py

similar physiological states, regardless of the patient. Additionally, the latent encoder $g_{\theta_g^e}^e$ and the recurrent predictor $g_{\theta_g^r}^r$ are encouraged to map temporally consecutive observations to spatially close latent coordinates, since it is expected that consecutive EHRs do not exhibit drastic changes, leading to smooth patient trajectories through the latent space. A neural decoder network:

$$d_{\theta_d} : \mathbb{R}^2 \to \mathbb{R}^D \tag{5.32}$$

reconstructs the original EHR features from the latent representation:

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

In this way, the decoder learns to disentangle the latent coordinates in $\hat{\boldsymbol{z}}_t$ based on ground truth future EHRs $\boldsymbol{\mu}_t$. The module is trained using a MSE supervised loss:

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

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.

5.2.4 Combining Infection and Acute Change Signals

The complete LDM, shown in Figure 5.4, is trained jointly by combining the previously introduced Infection Indicator Module f_{θ_f} and the SOFA prediction module g_{θ_g}. The output of these modules yield the components $\hat{O}_t$, 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:

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

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

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

with radius r a hyper-parameter controlling the temporal radius of the smoothing window, and α a learnable decay parameter controlling the length and shape of the smoothing kernel. To handle the sequence boundaries, it is assumed $x_{t-\tau} = 0$ 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. In reality, sepsis is not only a single time-point of onset but a physiological state that lasts some amount of time.
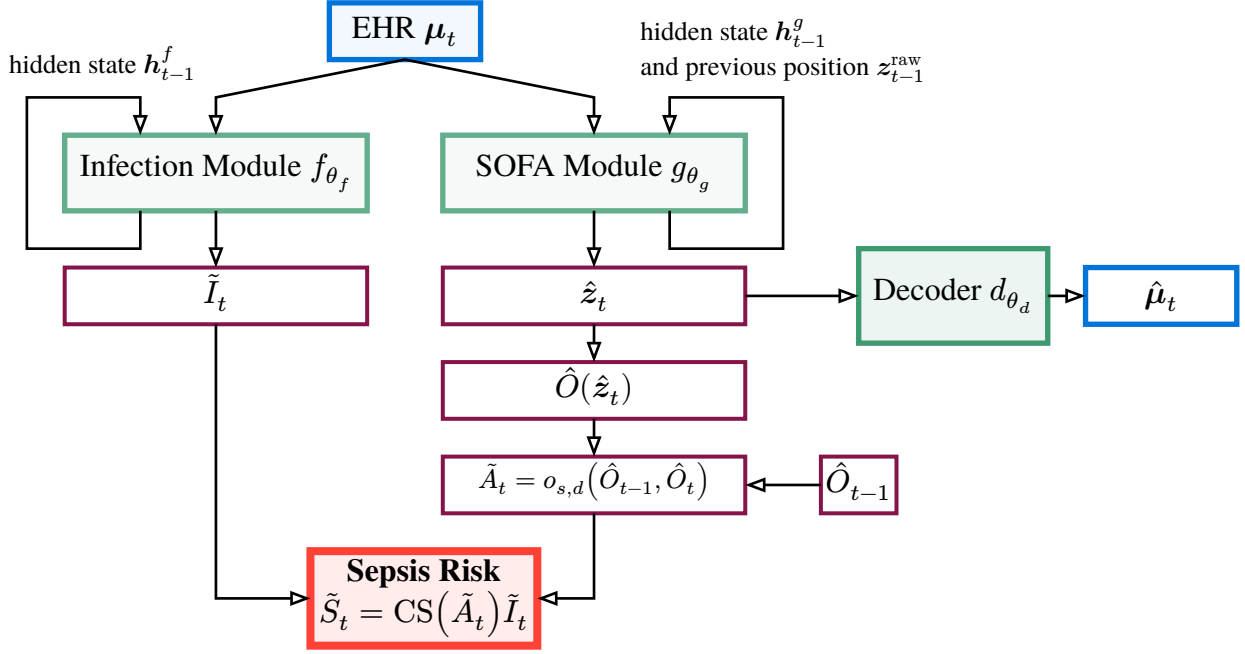
Figure 5.4: Complete LDM architecture with three main components. The Infection Module f_{θ_f} and SOFA Module g_{θ_g} process EHR data μ_t through recurrent networks to estimate infection level $\tilde{I}_t$ and latent coordinates $\hat{z}_t$ respectively. The latent coordinates map to organ failure $\hat{O}_t$, 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_{θ_d} reconstructs EHR features μ_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 Primary and Auxiliary Losses

Besides the losses already presented, multiple auxiliary losses are used to guide the training process 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:

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

using the BCE. Although the heuristic risk score $\tilde{S}_t$ is constructed from infection and organ dysfunction estimates, the primary sepsis loss $\mathcal{L}_{\text{sepsis}}$ is required to align their interaction with true clinical outcomes. This should help to compensate for modeling approximations and measuring uncertainties, while enabling the joint end-to-end optimization of all components for the actual prediction objective.

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

$$\mathcal{L}_{\text{spread}} = -\log\big(\det\big(\text{Cov}\big(\hat{\boldsymbol{Z}}\big)\big)\big) \tag{5.38}$$

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

The loss is minimized when $\det\big(\text{Cov}\big(\hat{\boldsymbol{Z}}\big)\big)$ of the latent dimensions β and σ is increased. This quantity is known as the *generalized variance* [64], and roughly measures the volume occupied by the distribution, where it increases when sampled points spread more. Consequently, $\mathcal{L}_{\text{spread}}$ encourages a larger spread in the latent space and prevents collapse to a narrow region.

Latent Space Regularization

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

$$\mathcal{L}_{\text{boundary}} = \text{ReLU}(f - \text{sigmoid}(\boldsymbol{z}_t^{\text{raw}})) + \text{ReLU}(\text{sigmoid}(\boldsymbol{z}_t^{\text{raw}}) - (1 - f)) \tag{5.39}$$

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. The Rectified Linear Unit (ReLU) activation function is defined as:

$$\text{ReLU}(x) = \max(x, 0) \tag{5.40}$$

and therefore nonzero only for positive inputs.

5.3.1 Combined Objective

The complete LDM[3] is trained jointly by optimizing all components via the combined weighted total loss:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{sepsis}}\mathcal{L}_{\text{sepsis}} + \lambda_{\text{inf}}\mathcal{L}_{\text{inf}} \qquad + \lambda_{\text{sofa}}\mathcal{L}_{\text{sofa}} + \tag{5.41}$$

$$\lambda_{\text{dec}}\mathcal{L}_{\text{dec}} + \lambda_{\text{spread}}\mathcal{L}_{\text{spread}} + \lambda_{\text{boundary}}\mathcal{L}_{\text{boundary}} \tag{5.41}$$

The loss weights λ 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 λ and other hyperparameters are reported in Section 6.2. Table 5.2 provides an overview of all loss components, their purpose, and the modules they supervise.

---

[3]Implementation of the LDM components is available at https://github.com/unartig/sepsis_osc/tree/main/src/sepsis_osc/ldm

Table 5.2: Overview of loss components in the LDM training objective.

Loss	Type	Purpose	Supervises
$\mathcal{L}_{\text{sepsis}}$	BCE	Primary sepsis prediction	$f_{\theta_f}, g_{\theta_g^e}^e, g_{\theta_g^r}^r$
$\mathcal{L}_{\text{sofa}}$	Weighted MSE	SOFA estimation	$g_{\theta_g^e}^e, g_{\theta_g^r}^r$
$\mathcal{L}_{\text{inf}}$	BCE	Infection indicator	f_{θ_f}
$\mathcal{L}_{\text{spread}}$	Covariance	Latent diversity	$g_{\theta_g^e}^e, g_{\theta_g^r}^r$
$\mathcal{L}_{\text{boundary}}$	Positional	Latent Space	$g_{\theta_g^e}^e, g_{\theta_g^r}^r$
$\mathcal{L}_{\text{dec}}$	MSE	Latent semantics	$d_{\theta_d}, (g_{\theta_g^e}^e, g_{\theta_g^r}^r)$

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 μ_1 is processed by both the infection indicator module and the latent encoder. The infection indicator f_{θ_f} produces an initial infection risk estimate $\tilde{I}_1$ and hidden state h_1^f. The latent encoder $g_{\theta_g^e}^e$ maps the μ_1 to the initial latent coordinates $\hat{z}_1$ and initial hidden state h_1^g. Deriving the synchronicity measure $s^1(\hat{z}_t)$ 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_1^f and h_1^g are saved to enable temporal continuity in subsequent predictions.

Triggered either by newly arriving measurements or by regular hourly intervals, the system performs sequential updates. Updated EHRs μ_t are processed by the recurrent modules f_{θ_f} and $g_{\theta_g^r}^r$ 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 $\hat{z}_{1:t}$ 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 to understand the contributing factors, whether the risk stems primarily from suspected infection, acute organ deterioration, or both. Additionally, the latent trajectory $s^1(\hat{z}_{1:t})$ through the DNM parameter space provides interpretable visualization of the patients physiological evolution over time.

5.5 Assessing the Prediction Performance

In order to quantitatively assess the sepsis prediction performance theoretically grounded metrics will be introduced. Predicting whether a patient will develop sepsis is a binary classification problem based on the continuous estimated 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:

$$\delta\left(\tilde{S}_t\right) = \mathbb{I}\left(\tilde{S}_t > \tau\right) \tag{5.42}$$

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

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

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

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5.43}$$

and the False Positive Rate (FPR):

$$\text{FPR} = \frac{\text{FP}}{\text{TP} + \text{FN}} \tag{5.44}$$

Sweeping the decision boundary τ 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 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 as a single number, the area under the curve, called AUROC, where larger values ≤ 1 indicate better prediction performance.

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

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5.45}$$

against the *recall*:

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5.46}$$

creating the *precision-recall curve* or PRC, where an optimal predictor hugs the top right. This curve can also be summarized by its area to the AUPRC metric, where larger values indicate better performance [65].

Traditionally, the AUPRC as been 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 [66]. Together the AUROC and AUPRC are the commonly reported performance metrics used in the sepsis prediction literature [5] and will be used to compare the performance between the LDM and a baseline approach.

5.6 Summary

This chapter introduced the proposed model for short-term and interpretable online 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, this 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.

In the next Chapter 6, an experiment where the LDM is trained using a widely used data-source ICU is presented. Therefore, it introduces the exact cohort definitions, data preprocessing 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, and generated latent trajectories are analyzed qualitatively.

6

Experiment

To evaluate whether embedding the DNM (see Chapter 4) improves short-term sepsis prediction, the LDM (see Chapter 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 ends 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 MIMIC-IV database (version 2.3) [38], which contains EHR information capturing day-to-day clinical routines, including patient measurements, orders, diagnoses, procedures, treatments, and free-text clinical notes. All included EHR data were recorded at Beth Israel Deaconess Medical Center in Boston, America between 2008 and 2022. Every part of the data has been anonymized and is publicly available to support research in electronic healthcare applications, with special focus on intensive care. Due to the individual particularities in data collection, applications trained on single datasets are known to have limited generalization to other datasources and real-world settings. Yet, MIMIC databases tend to be more generalizable than others and remain the default open-data resource for developing sepsis prediction systems [4, 67].

To derive a cohort from raw data and preprocess clinical features, the YAIB framework is employed [40]. 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 and feature choices to the data preprocessing has been adopted from their methodology [40] to enable direct comparison of prediction results.

Their definition of sepsis closely follows the Sepsis-3 criteria [12]:

"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 ≥ 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 ≤24 hours after antibiotic initiation or if antibiotics were started ≤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 ≥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"

6.1.1 Cohort Selection

The cohort includes all adult patients (age at admission ≥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 ≥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 of $N = 63,425$ patients. The selection process with corresponding exclusion numbers is shown in Figure 6.1.
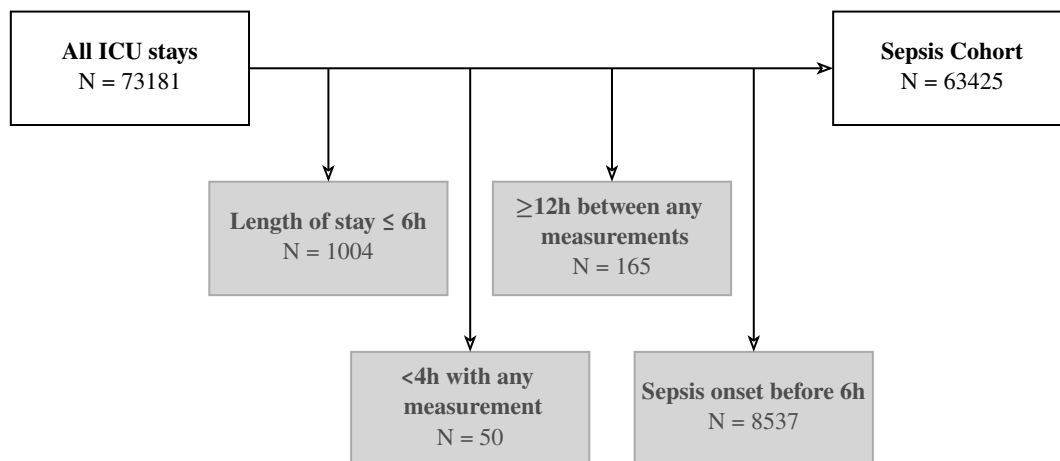
Figure 6.1: Cohort selection and exclusion process.

6.1.2 Cohort Characteristics

Table 6.1 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 with 25th-75th percentile interquantile range of 8–34 hours.

Both groups were similar with respect to demographic characteristics, including age (median 65 years), sex distribution (approximately 55% male), and weight at admission (median 77.6 kg). The majority of 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 6.1: Characteristics and demographics of the cohort. Numerical variables are summarized by *median [Interquantile Range 25th - 75th percentile]* and numerical variables by incidence (%)

CHARACTERISTIC	ALL PATIENTS	SEP-3 POSITIVE	SEP-3 NEGATIVE
Demographics			
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)
Clinical Outcomes			
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)	-
Ethnicity			
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)
Admission Type			
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 Feature Choice and Labeling

To enable direct comparison of results with the benchmark [40], their feature set is adopted, which has been derived in collaboration with clinical experts and includes only widely available clinical markers. Each EHR includes 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 A.3. provides a complete list of all features with their value ranges, units of measurement and clinical descriptions.

Primary prediction target is to detect the sepsis onset time within six hours. This means, that the positive label for a Sepsis-3 onset is stretched for six the hours preceding and proceeding the actual onset, creating a time window of 12 hours. For this work, the target variables additionally include the SOFA-score and a SI label for the auxiliary modules, in contrast to YAIB where only the Sepsis-3 label is used (see Section 5.1).

6.1.4 Preprocessing

Data preprocessing involved 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, effectively doubling the number of input features.

6.2 Implementation Details

Implementation of the LDM was done in the JAX [55] based Equinox framework [68]. Following sections present the implementation details for each module along with their respective parameter counts.

The infection indicator module f_{θ_f} consists of 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 A.4. of the Appendix A.4).

The latent encoder architecture $g_{\theta_g^e}^e$, illustrated in Figure 6.2/A, implements a gated attention mechanism with residual processing. Incoming samples of $\boldsymbol{\mu}_t$ 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 along the feature dimension [69]:

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

where γ and β are learnable parameters, with the same dimensionality as x. The normalization is followed by the Gaussian Error Linear Unit (GELU) activation [70], and a linear transformation with residual connections. With GELU:

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

where $\Phi(x)$ is the cumulative distribution function for Gaussian distribution. Finally, the last hidden state passes through another GELU-activated layer before being projected into the two outputs h_1 and z_1^{raw}. In total the encoder has $19,350$ parameter.

The rollout module $g_{\theta_g^r}^r$ 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 up 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 and is not performed in this work.
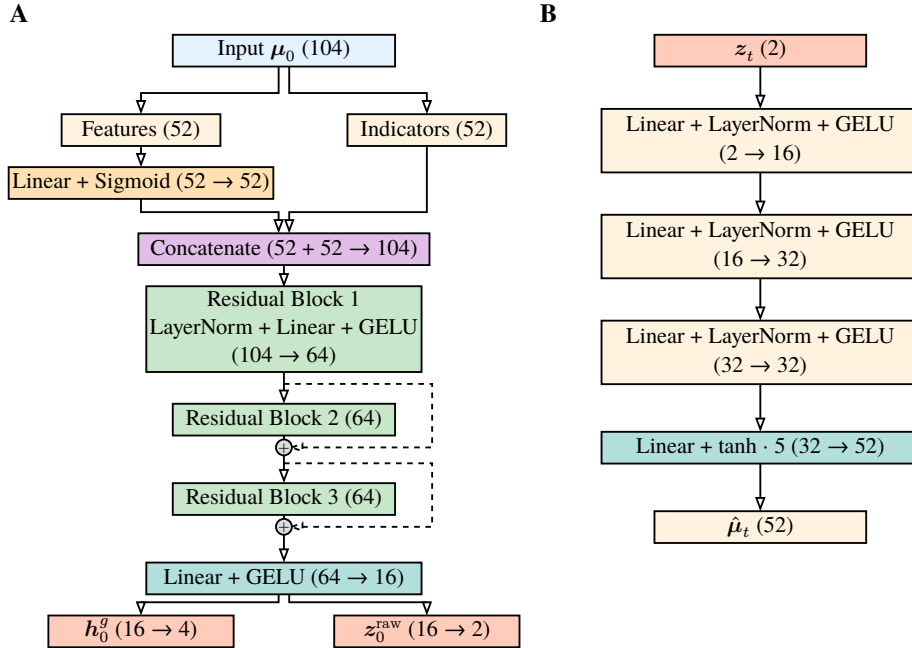
Figure 6.2: **A** Shows the initial latent encoder $g_{\theta_g^e}$ architecture with feature gating and residual connections. Dashed arrows indicate residual skip connections. **B** Shows the decoder d_θ^d architecture, reconstructing only the features but not the imputation indicators.

The decoder d_{θ_d}, illustrated in Figure 6.2/B, 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, followed by layer-normalization, and a

final `tanh` activation and scaling by a factor of 5 on the output to structure it according to the normalized input data with mean zero and unit variance. In total the decoder has $3,524$ learnable parameter.

6.2.1 Training Details

The cohort was partitioned at the patient level using a stratified split with an 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 infeasible for the scope of this study. The primary goal was to demonstrate feasibility and interpretability rather than achieving state-of-the-art performance. Table 6.2 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 [71]. In the end, the magnitude of the loss-gradients matters, not the raw numerical values.

All modules were jointly optimized using the AdamW optimizer [72] 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 5×10^{-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 60×100 grid over $\beta \in [0.4\pi, 0.7\pi]$ and $\sigma \in [0.0, 1.5]$, with differentiable lookup using a 3×3 neighborhood softmax interpolation ($K = 9$).

Training was carried out on a consumer laptop graphics card (NVIDIA RTX 500 Ada Generation with 4 GB of VRAM) on 32-bit floating-point precision. Training lasted a maximum of 1000 epochs, and 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 6.2: 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
Hyperparameter			
λ_{sepsis}	600.0	Weight of $\mathcal{L}_{\text{sepsis}}$	Equation (5.41)
λ_{sofa}	2×10^3	Weight of $\mathcal{L}_{\text{sofa}}$	Equation (5.41)
λ_{inf}	1.0	Weight of $\mathcal{L}_{\text{inf}}$	Equation (5.41)
λ_{spread}	6×10^{-3}	Weight of $\mathcal{L}_{\text{spread}}$	Equation (5.41)
$\lambda_{\text{boundary}}$	30.0	Weight of $\mathcal{L}_{\text{boundary}}$	Equation (5.41)
λ_{dec}	5.0	Weight of $\mathcal{L}_{\text{dec}}$	Equation (5.41)
τ	12	Radius of causal smoothing	Equation (5.36)
k	3	Side length of the latent-lookup kernel	Equation (5.29) Equation (5.31)
f	0.1	Fraction before $\mathcal{L}_{\text{boundary}}$ triggers	Equation (5.39)
Learnable Parameter			
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)
α	0.7	Causal smoothing decay	Equation (5.36)

6.3 Results

The ollowing 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 6.3 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, declines 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 the 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 representations, namely the alignment
of the latent space with organ dysfunction, as well as SI.
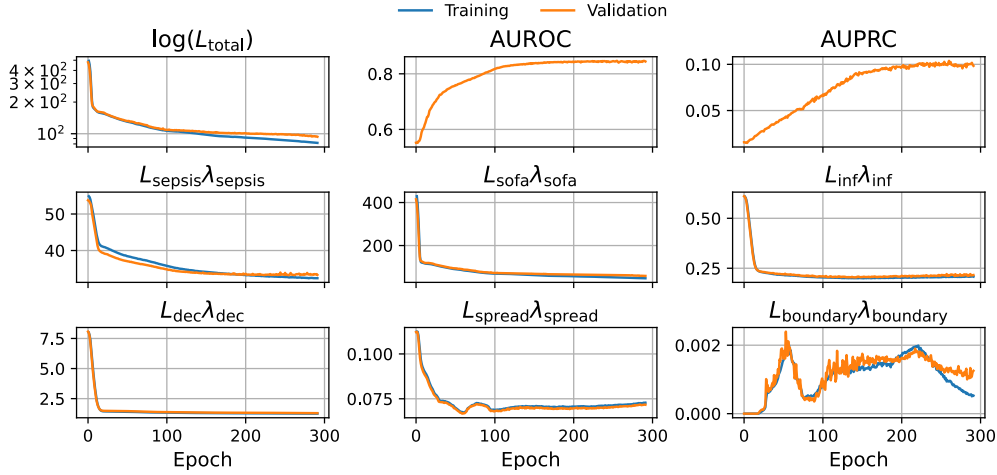
Figure 6.3: Training and validation curves of the LDM showing the evolution of the total loss, task
metrics (AUROC, AUPRC), and all individual loss components. The plots illustrate
stable multi-objective convergence, early alignment of infection f_{θ_f} and SOFA g_{θ_g}
submodules, and gradual refinement of the final sepsis risk prediction without signs of
overfitting.

Overall, the decoder reconstruction loss $\mathcal{L}_{\text{dec}}$ exhibits a typical 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 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 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 his-
tories. 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.

6.3.2 Quantitative Results

This training run was stopped early 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 6.4 by sweeping the decision threshold τ from
Equation (5.42). Here, the ROC curve exhibits the expected shape, bending toward the top-
left corner and away from the chance level curve shown as the dashed line. An AUROC of
0.8469 shows 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 τ), 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).
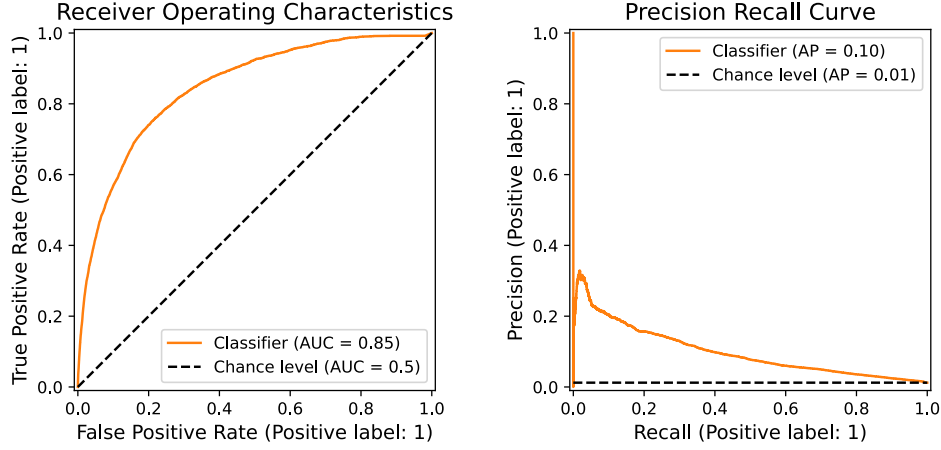
Figure 6.4: Receiver Operating Characteristic and precision-recall curves generated by sweeping the detection threshold τ from Equation (5.42).

To put these results into perspective, Table 6.3 compares the LDM performance to baseline models trained in YAIB [40] on the same task, data-source, cohort definition and train-test split. The baseline models include classical ML and DL approaches, with hyperparameters tuned via Bayesian optimization, 30 iterations for ML models, 50 for DL models. Baseline DL models, including GRU, LSTM, Temporal Convolutional Networks, and Transformer outperform the classical ML baselines, including Regularized Logistic Regression and Light Gradient Boosted Machines, systematically. The LDM outperforms all baselines in both metrics.

Table 6.3: Performance Comparison against the mean performance of the baseline models trained in YAIB [40], in terms of AUROC $\times$ 100 ($\uparrow$, higher is better) and AUPRC $\times$ 100 ($\uparrow$). Best performances in **bold**.

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

In Figure 6.5 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 of the 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 high 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 (ΔSOFA, 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 any 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 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.
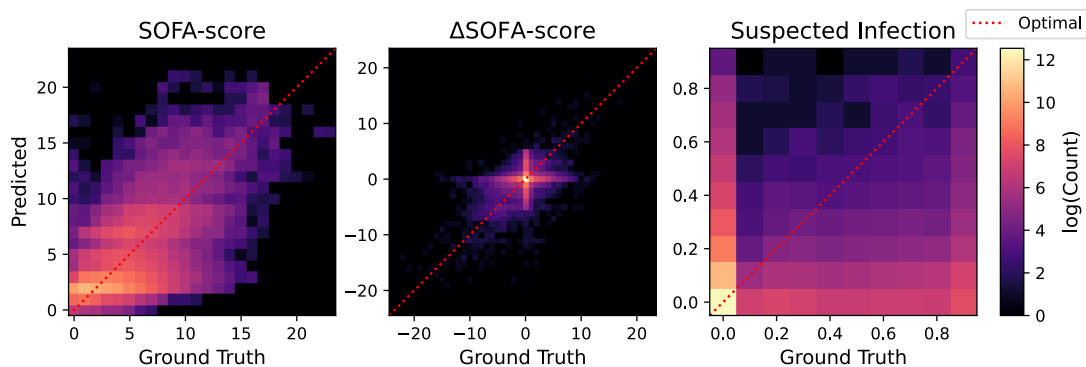

Figure 6.5: Density plots comparing ground truth and predicted values for SOFA-score, immediate SOFA change (ΔSOFA), and SI. The model captures the overall SOFA severity distribution and its temporal changes, while infection predictions show reasonable separation between low- and high-probability states. Color indicates log sample density.

6.3.3 Qualitative Results

This section investigates, whether the model is able to create plausible patient trajectories inside DNM parameter space. First of all, Figure 6.6/A shows predicted sample density of $\hat{z}$ layered over the latent space, where both the smooth low-desynchronized between β values 0.4 and 0.5 and the more dynamic highly-desynchronized area between β 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 $\hat{z}_{\text{center}} \approx (\beta = 0.58, \sigma = 0.98)$, from there, individual trajectories spread out into all directions, mostly ending up in the low-desynchronized.

In Figure 6.6/B, the predicted latent positions colored by the ground truth SOFA-scores are layered over the latent space. Since brighter colors for predicted $s^1(\hat{z})$ 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.
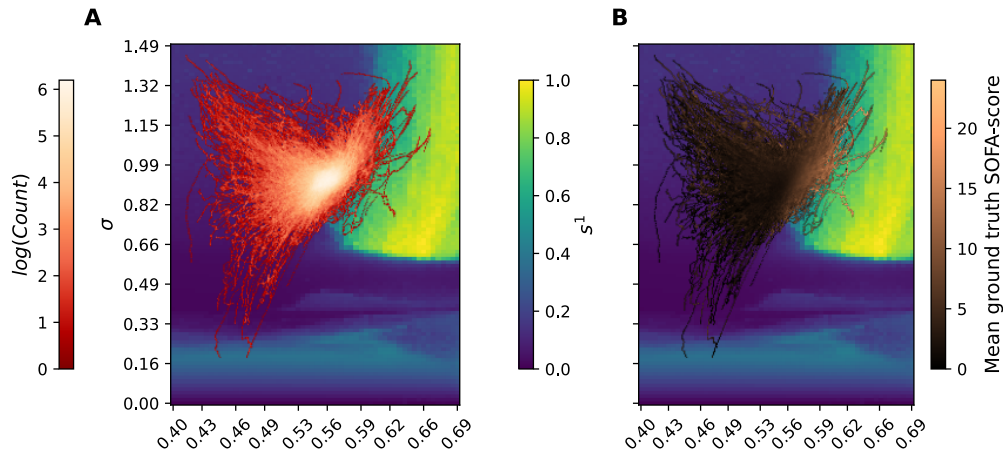
Figure 6.6: **A** shows the distribution of patient EHR encoded as latent points $\hat{z} = (\hat{z}_\beta, \hat{z}_\sigma)$. The latent space is colored with the values of the normalized desynchronization metric $s^1(\hat{z})$, where brighter values indicate larger desynchronization. The point distribution is colored by density, with brighter values having greater density. **B** shows the same latent space, but the overlay points are colored by the mean ground truth SOFA-score of all points for each coordinate, here the desired outcome is that the color gradients align.

Moving from the system-wide behavior to three individual patient trajectories, shown in Figure 6.7. These hand-picked examples illustrate possible physiological progressions and how each is represented inside the DNM parameter space. The figure includes three patients, showing predicted and ground truth SOFA-score evolution on the right side, with the left side showing corresponding latent trajectories colored by the ground truth values.

Patient 1 (Figure 6.7, cyan) at ICU admission $t = 0$, has a relatively high SOFA-score of 12, and the organ system deteriorates further, 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 10 score points, yet it picks up the trend of deterioration, but overshoots and arrives at a SOFA-score of 20. This progression is mirrored 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 (Figure 6.7, purple) 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 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 (Figure 6.7, pink) 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.
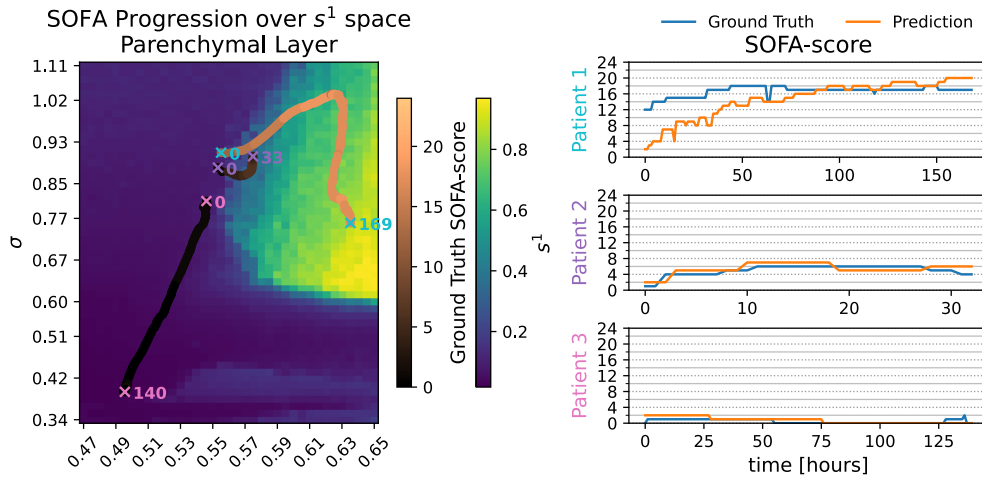
Figure 6.7: Left: Three patient trajectories plotted in the latent DNM parameter space (β, σ), with their curves colored by ground truth SOFA-scores in brownish colors. Timestamps mark trajectory beginning and end. The background heatmap shows s^1 values (right colorbar), with darker regions indicating 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.

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 6.5 and heterogeneity of trajectories in Figure 6.6/B.

7
Discussion

This work introduced the LDM, a novel architecture that embeds the DNM as an interpretable latent space for short-term sepsis prediction from 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:

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?

Experimental results provide encouraging answers to both questions. Regarding usability, it was demonstrated that DNM parameters (β, σ) 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%, 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. While other methodologies can adopt the sepsis label decomposition, the additional visual information of the latent trajectory is unique to the LDM.

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 Chapter 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 [37, 39]. 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 6.6) 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 6.7). This multi-scale interpretability aligns with clinician preferences identified in [30]. 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_{θ_f}, g_{θ_g} and d_{θ_d}, it provides the necessary flexibility to improve on individual aspects of the system in a straight forward manner.

7.3 Limitations and Challenges

While the experimental results are encouraging, several important limitations need careful consideration. As discussed in Section 4.4, the DNM faces inherent limitations. Parameters like β (biological age) and σ (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.

Missing ablation studies to understand the impact of each of the LDM components have not been performed, but are crucial to understand the role and importance of each. This is especially critical for the DNM, even though it provides a physiologically motivated latent space, it remains unknown whether it is solely the model built around the DNM parameter space that drives prediction performance or how much the inductive bias of the DNM contributes. Furthermore, the LDM currently does not implement the DNM dynamical's system directly since it does not solve the coupled differential equations for each latent position individually. Instead, it learns to position patients in a parameter space abstracted from the DNM, with the intent (encouraged by loss functions) that these positions correlate meaningfully with true physiological states. Ultimately, 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. Direct mechanistic interpretability cannot be claimed, a more honest characterization might be 'physiologically-motivated dimensionality reduction'.

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 ($\pm 1\%$ 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 high-dimensional patient state, represented by the EHR, to the low dimensional parameter space of the DNM. This ambiguity offers infinitely many mappings, while some 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, this variability in solutions should be acknowledged rather than treating a single trained model as definitive.

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 that the prediction has to move from plain distinction between septic and non-septic to sensitive estimates of 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 [73].

7.4 Future Directions

All experiments relied on the MIMIC-IV exclusively. While this enables direct comparison with YAIB benchmarks, it limits generalizability claims. As noted in Chapter 3, applications trained on single datasources often do not generalize well to other datasources or real-world settings. This is why, 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 transferable. Another point of interest is the particularity of the DNM parameter space. By replacing it with other biologically (un-)motivated spaces, one could test if the DNM provides especially informative priors to the LDM or not.

Currently, 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_{θ_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 be used to deduce practical patient-individual treatment?

Furthermore, the encoder $g_{\theta_g^e}^e$ 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.

The 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 $\hat{z}_t = (\beta_t, \sigma_t)$, solve this system numerically to obtain $s^1(\hat{z}_t)$. Modern differentiable ODE-solvers, such as diffrax [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 mechanistic and more nuanced interpretability.

This work focused on online prediction, where risk estimates update continuously as new measurements arrive. As noted in Chapter 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 additions. 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.

Lastly, some general concerns, not specific to the LDM. Current sepsis prediction implementations 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.

8
Conclusion

This work has motivated and demonstrated that hybrid modeling, integrating biologically grounded dynamics with data-driven learning, can overcome limitations of purely mechanistic and purely black-box approaches to sepsis prediction. The starting point was the DNM [8, 9], a functional model of coupled oscillators, representing organ and immune cell populations that can describe both healthy and pathological states related to sepsis. Previously untested on clinical data, in this work it was validated through the LDM framework, a novel neural architecture that embeds the DNM within a DL architecture for online sepsis prediction.

The LDM decomposes binary sepsis prediction into its underlying Sepsis-3 components: suspected infection and acute organ dysfunction. Crucially, it represents patient organ system states as trajectories through the DNMs parameter space, modulating the degree of frequency synchronization, serving as a proxy for acute organ failure. This design enables the model to learn mappings from electronic health records to physiologically meaningful latent coordinates while preserving the structured inductive bias of the DNM.

Trained and evaluated retrospectively on real-world patient data, the LDM achieved an AUROC of 84.69 and AUPRC of 9.87, outperforming baseline models by at least 1.9% AUROC and 0.77% AUPRC. Beyond improved predictive accuracy, the LDM provides interpretable intermediate outputs, in contrast to purely data-driven black-box models. Qualitative analysis of patient trajectories through the latent space demonstrated clinically plausible patterns of deterioration, recovery, and stability, with latent coordinates systematically aligning with physiological progressions.

This proof-of-concept study establishes hybrid physics-informed deep learning as a promising pathway for interpretable clinical decision support, offering an alternative to purely black-box models while maintaining competitive predictive performance. Finally, this work has acknowledged current limitations, such as external validation and the lack of systematic latent-space analyses, but also provides directions for future research in terms of model improvements and extensions.

A Appendix

A.1 SOFA - Details

Table A.1.: Components of the SOFA-score definition [24].

CATEGORY	INDICATOR	1	2	3	4
Respiration	PaO$_2$/FiO$_2$ [mmHg]	< 400	< 300	< 200	< 100
	Mechanical Ventilation			yes	yes
Coagulation	Platelets [$\times \frac{10^3}{mm^3}$]	< 150	< 100	< 50	< 20
Liver	Bilirubin [$\frac{mg}{dl}$]	1.2-1.9	2.0-5.9	6.0-11.9	> 12.0
Cardiovascular[4]	MAP [mmHg]	< 70			
	or Dopamine		≤ 5	> 5	> 15
	or Dobutamine		any dose		
	or Epinephrine			≤ 0.1	> 0.1
	or Noepinephrine			≤ 0.1	> 0.1
Central Nervous System	Glasgow Coma Score	13-14	10-12	6-9	< 6
Renal	Creatinine [$\frac{mg}{dl}$]	1.2-1.9	2.0-3.4	3.5-4.9	> 5.0
	or Urine Output [$\frac{ml}{day}$]			< 500	< 200

---

[4]Adrenergica agents administered for at least 1h (doses given are in [µg/kg · min]

A.2 Lie formulation of the Dynamic Network Model

The idea from the reformulation originated from an encounter with [74]. Let $SO(2)$ denote the special orthogonal group of 2D rotations [75]. Each element $R \in SO(2)$ can be represented as:

$$R(\varphi) = \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix} \tag{1.1}$$

With rotation operators:

$$R_i^1 := R(\varphi_i^1), \quad R_i^2 := R(\varphi_i^2) \tag{1.2}$$

$$R_\alpha := R(\alpha), \quad R_\beta := R(\beta) \tag{1.2}$$

The inverse rotation is $R^{-1}(\varphi) = R(-\varphi)$, and the relative rotation between two phases is:

$$R_{ij} := (R_i)^{-1} R_j = R(\varphi_j - \varphi_i) \tag{1.3}$$

The sine of a rotation difference can be extracted as:

$$\sin(\varphi_i - \varphi_j + \gamma) = \left[(R(\varphi_i - \varphi_j)R(\gamma)) \right]_{2,1} = \left[R_{ij} R_\gamma \right]_{2,1} \tag{1.4}$$

where $[\cdot]_{2,1}$ denotes the (2,1) matrix element.

The Lie group formulation of the DNM is:

$$\dot{\varphi}_i^1 = \omega^1 - \frac{1}{N} \sum_{j=1}^N \left\{ (a_{ij}^1 + \kappa_{ij}^1) \left[(R_{ij}^1 R_{\alpha^{11}}) \right]_{2,1} \right\} - \sigma \left[\left(R_i^1 (R_i^2)^{-1} R_{\alpha^{12}} \right) \right]_{2,1} \tag{1.5}$$

$$\dot{\kappa}_{ij}^1 = -\varepsilon^1 \left(\kappa_{ij}^1 + \left[\left(R_{ij}^1 R_\beta^{-1} \right) \right]_{2,1} \right) \tag{1.6}$$

$$\dot{\varphi}_i^2 = \omega^2 - \cdot \frac{1}{N} \sum_{j=1}^N \left\{ \kappa_{ij}^2 \left[(R_{ij}^2 R_{\alpha^{22}}) \right]_{2,1} \right\} - \sigma \left[\left(R_i^2 (R_i^1)^{-1} R_{\alpha^{21}} \right) \right]_{2,1} \tag{1.7}$$

$$\dot{\kappa}_{ij}^2 = -\varepsilon^2 \left(\kappa_{ij}^2 + \left[\left(R_{ij}^2 R_\beta^{-1} \right) \right]_{2,1} \right) \tag{1.8}$$

Implementation[5] is based in jaxlie [76]. Theoretical stability benefits are mitigated by elevated construction costs of the rotation matrices and do not translate to better integration performance.

---

[5]available at https://github.com/unartig/sepsis_osc/blob/main/src/sepsis_osc/dnm/lie_dnm.py

A.3 Input Concepts

The following two tables list the static and dynamic input features used by the YAIB framework [40]. The column "ricu-name" refers to the R package `ricu` which YAIB is built on top of [77].

Table A.2.: 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 A.3.: 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 A.4.: Detailed parameter count of the Latent Dynamics Model modules.

NAME	SHAPE	COUNT	NAME	SHAPE	COUNT
Infection Module			**SOFA Module (Encoder)**		
GRU-Cell weight_ih	(96, 104)	9,984	gating weight	(52, 52)	2,704
GRU-Cell weight_hh	(96, 32)	3,072	norm1 weight	(104,)	104
GRU-Cell bias	(96,)	96	norm1 bias	(104,)	104
GRU-Cell bias_n	(32,)	32	linear1 weight	(64, 104)	6,656
h0	(32,)	32	linear1 bias	(64,)	64
Projection weight	(1, 32)	32	norm2 weight	(64,)	64
Projection bias	(1,)	1	norm2 bias	(64,)	64
	Total:	**13,249**	linear2 weight	(64, 64)	4,096
Decoder Module			linear2 bias	(64,)	64
linear1 weight	(16, 2)	32	norm3 weight	(64,)	64
linear1 bias	(16,)	16	norm3 bias	(64,)	64
norm1 weight	(16,)	16	linear3 weight	(64, 64)	4,096
norm1 bias	(16,)	16	linear3 bias	(64,)	64
linear2 weight	(32, 16)	512	linear4 weight	(16, 64)	1,024
linear2 bias	(32,)	32	linear4 bias	(16,)	16
norm2 weight	(32,)	32	z weight	(2, 16)	32
norm2 bias	(32,)	32	z bias	(2,)	2
linear3 weight	(32, 32)	1,024	h weight	(4, 16)	64
linear3 bias	(32,)	32	h bias	(4,)	4
norm3 weight	(32,)	32		**Total:**	**19,350**
norm3 bias	(32,)	32	**SOFA Module (RNN)**		
linear4 weight	(52, 32)	1,664	GRU-Cell weight_ih	(12, 106)	1,272
linear4 bias	(52,)	52	GRU-Cell weight_hh	(12, 4)	48
	Total:	**3,524**	GRU-Cell bias	(12,)	12
General			GRU-Cell bias_n	(4,)	4
lookup_temp	(1,)	1	Proj weight	(2, 4)	8
causal_decay	(1,)	1		**Total:**	**1,344**
detection_threshold	(1,)	1			
detection_sharpness	(1,)	1			
	Total:	**4**			

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] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] D. H. P. Directors, "Overview of Surface Epithelium." [Online]. Available: https://digitalhistology.org/tissues/epithelium/lining-and-covering/overview/overview-1/

[16] D. H. P. Directors, "Stroma and Parenchyma." [Online]. Available: https://digitalhistology.org/organs-systems/general-concepts/stroma-and-parenchyma/stroma-and-parenchyma-1/

[17] S. Fischer and E. Deindl, "State of the Art of Innate Immunity - An Overview," *Cells*, vol. 11, p. 2705, 2022, doi: 10.3390/cells11172705.

[18] D. Jarczak, S. Kluge, and A. Nierhaus, "Sepsis—Pathophysiology and Therapeutic Concepts," *Frontiers in Medicine*, 2021, doi: 10.3389/fmed.2021.628302.

[19] 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.

[20] R. V. House and J. Descotes, *Cytokines in Human Health: Immunotoxicology , Pathology , and Therapeutic Applications*, 1st ed. in Methods in Pharmacology and Toxicology. Totowa , N.J.: Humana Press, 2007.

[21] 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.

[22] C. Garlanda, I. Di Ceglie, and S. Jaillon, "IL-1 family cytokines in inflammation and immunity," *Cellular & Molecular Immunology*, vol. 22, no. 11, pp. 1345–1362, 2025, doi: 10.1038/s41423-025-01358-8.

[23] 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.

[24] 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.

[25] 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.

[26] N. Bennett, D. Plečko, and I.-F. Ukor, "Sepsis 3 label — sep3." [Online]. Available: https://eth-mds.github.io/ricu/reference/label_sep3.html

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] A. Schuurman, P. Sloot, W. Wiersinga, and et al., "Embracing complexity in sepsis," *Critcal Care*, vol. 27, no. 1, p. 102, 2023, doi: 10.1186/s13054-023-04374-0.

[32] 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

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] A. E. W. Johnson and others, "MIMIC-IV , a freely accessible electronic health record dataset," *Scientific Data*, vol. 10, no. 1, p. 1, Jan. 2023, doi: 10.1038/s41597-022-01899-x.

[39] 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.

[40] 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.

[41] 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.

[42] P. C. Ivanov, "The New Field of Network Physiology: Building the Human Physiolome," *Frontiers in Network Physiology*, 2021, doi: 10.3389/fnetp.2021.711778.

[43] 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.

[44] 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.

[45] K. Lehnertz, T. Rings, and T. Bröhl, "Time in Brain: How Biological Rhythms Impact on EEG Signals and on EEG-Derived Brain Networks," *Frontiers Network Physiology*, vol. 1, p. 755016, Sept. 2021, doi: 10.3389/fnetp.2021.755016.

[46] 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.

[47] 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.

[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," *Physical Review Letters*, 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." [Online]. Available: https://arxiv.org/abs/2202.02435

[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] 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

[60] 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

[61] F. Mentzer, D. Minnen, E. Agustsson, and M. Tschannen, "Finite Scalar Quantization: VQ-VAE Made Simple." [Online]. Available: https://arxiv.org/abs/2309.15505

[62] D. Hafner, J. Pasukonis, J. Ba, and T. Lillicrap, "Mastering Diverse Domains through World Models." [Online]. Available: https://arxiv.org/abs/2301.04104

[63] 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

[64] 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.

[65] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge , MA : MIT Press, 2012.

[66] 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

[67] 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.

[68] P. Kidger and C. Garcia, "Equinox: neural networks in JAX via callable PyTrees and filtered transformations." [Online]. Available: https://arxiv.org/abs/2111.00254

[69] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer Normalization." [Online]. Available: https://arxiv.org/abs/1607.06450

[70] D. Hendrycks and K. Gimpel, "Gaussian Error Linear Units (GELUs)." [Online]. Available: https://arxiv.org/abs/1606.08415

[71] 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

[72] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization." [Online]. Available: https://arxiv.org/abs/1711.05101

[73] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On Calibration of Modern Neural Networks." [Online]. Available: https://arxiv.org/abs/1706.04599

[74] M. A. M. de Aguiar, "On the numerical integration of the multidimensional Kuramoto model." [Online]. Available: https://arxiv.org/abs/2306.05939

[75] K. Lynch and F. Park, *Modern Robotics: Mechanics , Planning , and Control*. Cambridge University Press, 2017. [Online]. Available: https://books.google.de/books?id=8uS3AQAACAAJ

[76] B. Yi, M. Lee, A. Kloss, R. Martín, and J. Bohg, "Differentiable Factor Graph Optimization for Learning Smoothers," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021.

[77] N. Bennett, D. Plečko, I.-F. Ukor, N. Meinshausen, and P. Bühlmann, "ricu: R's interface to intensive care data," *GigaScience*, vol. 12, p. giad41, 2023.