TU WIEN Informatics

Prädiktive temporale Entscheidungsmodellierung mit dynamischen Bayes-Netzwerken für frühzeitige klinische Diagnose

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Data Science

eingereicht von

Mariusz Nitecki, B.Sc.
Matrikelnummer 11711554

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Associate Prof. Dr. techn. Dipl.-Ing. Clemens Heitzinger

Wien, 10. Oktober 2025

Mariusz Nitecki	Clemens Heitzinger

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

TU Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

TU WIEN Informatics

Predictive Temporal Decision Modeling with Dynamic Bayesian Networks for Early Clinical Diagnosis

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Data Science

by

Mariusz Nitecki, B.Sc.
Registration Number 11711554

to the Faculty of Informatics

at the TU Wien

Advisor: Associate Prof. Dr. techn. Dipl.-Ing. Clemens Heitzinger

Vienna, October 10, 2025

Mariusz Nitecki	Clemens Heitzinger

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.at

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Erklärung zur Verfassung der Arbeit

Mariusz Nitecki, B.Sc.

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.
Ich erkläre weiters, dass ich mich generativer KI-Tools lediglich als Hilfsmittel bedient habe und in der vorliegenden Arbeit mein gestalterischer Einfluss überwiegt. Im Anhang „Übersicht verwendeter Hilfsmittel“ habe ich alle generativen KI-Tools gelistet, die verwendet wurden, und angegeben, wo und wie sie verwendet wurden. Für Textpassagen, die ohne substantielle Änderungen übernommen wurden, haben ich jeweils die von mir formulierten Eingaben (Prompts) und die verwendete IT-Anwendung mit ihrem Produktnamen und Versionsnummer/Datum angegeben.

Wien, 10. Oktober 2025

Mariusz Nitecki

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Kurzfassung

Eine frühzeitige Erkennung von Sepsis ist entscheidend, um rechtzeitige medizinische Maßnahmen zu gewährleisten und dem Patienten eine schnelle Genesung zu ermöglichen. Diese Arbeit untersucht, ob ein dynamisches Bayes-Netz (DBN) Sepsis früher erkennen kann als der etablierte Sequential Organ Failure Assessment Score (SOFA), und ob das Ausmaß signifikant ist. Zusätzlich wird ein Value-of-Information-Ansatz (VoI) angewandt, um den erwarteten Informationsgewinn jeder Laboruntersuchung zu berechnen und zu priorisieren. Anschließend wird geprüft, ob sich durch diese Strategie die Sepsiserkennung im DBN weiter beschleunigen und die diagnostische Unsicherheit verringern lässt.

Dafür wurde eine optimale DBN-Struktur mit Hilfe eines Hill-Climbing Verfahrens erstellt. Diese berücksichtigt wichtige medizinische Einschränkungen, um klinische Plausibilität zu gewährleisten. Die Parametrisierung erfolgte anhand von Daten von Intensivstationen aus der MIMIC-IV-Datenbank. Die Sepsisvorhersagen des DBN wurden anschließend mit denen des SOFA-Scores verglichen. Der VoI-Ansatz priorisiert Laborparameter nach ihrem Beitrag zur Reduktion diagnostischer Unsicherheit und erstellt damit für jede Patientin und jeden Patienten eine optimierte Messzeitlinie.

Die Ergbnisse dieser Arbeit zeigen, dass der DBN Ansatz eine Sepsis Diagnose durchschnittlich 4,2 Stunden früher stellt als der SOFA-Score. Die Anwendung der VoI Laborpriorisierungsmethode verbessert diesen zeitlichen Vorsprung auf 7,1 Stunden. Glecihzeitig nimmt die diagnostische Unsicherheit signifikant ab. Die Arbeit unterstreicht damit das Potenzial eines DBN und VoI Ansatzes für eine frühere und zuverlässigere Sepsisdiagnose.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Abstract

Early detection of sepsis is critical for effective clinical intervention and improving patient outcomes. This thesis investigates whether a Dynamic Bayesian Network (DBN) can identify sepsis earlier than the traditional Sequential Organ Failure Assessment (SOFA) score and analyses the potential lead time of such a new method. Additionally, a Valueof Information (VoI) approach is implemented for ordering laboratory tests. This method is then evaluated within the DBN framework and weather it can further accelerate sepsis detection and reduce diagnostic uncertainty.

To address these research questions, an optimal DBN structure is constructed with the help of data driven hill climbing search algorithm, which incorporates some medicaldomain constraints to ensure clinical plausibility. The DBN parameters are then trained on conditional probability distributions from patient data in intensive care units sourced from the MIMIC-IV database. The models sepsis predictions are compared to the SOFA scoring method. Furthermore, the VoI approach systematically orders laboratory tests by selecting those measurements providing the greatest reduction in diagnostic uncertainty,thus constructing an alternative optimized timeline for each patient.

The results demonstrate that the DBN identifies sepsis significantly earlier than the SOFA score, achieving a significant lead time advantage of approximately 4.2 hours. Incorporating the VoI-guided strategy further improves the lead time of predictions, increasing it to approximately 7.1 hours, while minimizing diagnostic uncertainty. These findings show a clinical potential of integrating DBNs and VoI methods for earlier sepsis diagnosis.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Contents

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

CHAPTER 1

Datasets

1.1 The MIMIC-IV Database

The Medical Information Mart for Intensive Care (MIMIC) project is a publicly accessible repository of electronic health record data collected at the Beth Israel Deaconess Medical Center in Boston, Massachusetts. The MIMIC-IV v3.1 dataset contains 364,627 unique patients, 546,028 hospital admissions, and 94,458 intensive care unit (ICU) stays recorded between 2008 and 2022. Additional admissions through 2022 have increased the dataset by approximately 65,000 patients, 115,000 admissions, and 21,000 ICU stays [1]. The database captures patient information such as demographics, admission and discharge details, laboratory measurements, and high frequency vital signs among many other essential information needed for treatment. All tables have standardized identifiers for patients, admissions and events.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

1.1.1 Data Modules

MIMIC-IV is organized into two complementary modules, `hosp` and `icu`, containing information about the stages of patient care. `hosp` contains data from the hospital's clinical information system, including patient demographics, admission and discharge details, intra-hospital transfers, pharmacy orders, laboratory results, microbiology cultures, and high-frequency vital-sign streams. Diagnoses and procedures were encoded into their respective International Classification of Diseases (ICD) codes. In contrast, `icu` captures information specific to intensive care stays documented in the MetaVision bedside system. Each ICU stay is linked to a set of event tables (a subset of those in `hosp`) but is limited to the ICU time frame. Both modules share standardized identifiers for patients, admissions, and events, allowing joins and longitudinal analyzes between modules [1].

Three main tables store relevant information about each hospital stay; `patients` for personal and demographic details, `admissions` for general hospital visits, and `icustays` for time spent in the intensive care unit. Medical events that contain necessary information to accurately predict sepsis are stored in specific tables depending on the type of data. Medication orders and records are found in `prescriptions`, `inputevents`, and `outputevents`. Lab test results are stored in `labevents`, and bedside observations are recorded in `chartevents`. Results from microbiology tests are found in `microbiologyevents`. Diagnoses and procedures, coded using ICD-9-CM or ICD-10-CM standards, are kept in the `diagnoses_icd` and `procedures_icd` tables [2].

1.1.2 Structure of the ICU Module

An overview of the row counts for key tables within the `icu` module of MIMIC-IV version 3.1 [1] is shown in Table 1.1. The `icu` module documents high resolution clinical and measured laboratory data for ICU stays. This includes bedside observations, laboratory measurements, medication administration, and procedures.

Table 1.1: Row counts for selected `icu` tables in MIMIC-IV v3.1. (as of July 2024)

Table	Entries
`chartevents`	432 997 491
`labevents`	158 374 764
`inputevents`	10 953 713
`outputevents`	5 359 395
`procedureevents`	808 706

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

The chartevents table is the largest table in the icu module. With over 430 million entries, it records time stamped clinical observations, include vital signs (such as heart rate, blood pressure, and respiratory rate), medication infusion rates, and clinical assessments like the Glasgow Coma Scale (GCS). Each observation is linked to an ICU stay through the hospital admission identifier hadm_id and time stamped by a unique charttime. Each observation and physiological measurement recorded in chartevents is identified by an itemid identification number [1].

Another key table, labevents, contains approximately 158 million laboratory test results from both ICU and non-ICU hospital stays. These include a wide range of biochemical, hematological, and microbiological observations. Similar to the chartevents table, all measurements are time stamped. Each result is referenced by hadm_id, and charttime, with the specific test indicated by a unique itemid. There are over 1,650 distinct laboratory tests cataloged in the accompanying d_labitems dictionary, covering measurements such as blood gas analyses, complete blood counts, etc. [1].

Intravenous fluids and medication administrations are recorded in the inputevents table. For each entry it stores start time, total dose, infusion rate, and medication label. The outputevents table documents patient outputs such as urine output, drain volumes, and other fluid losses. Finally, procedureevents lists bedside procedures (e.g. intubation, arterial-line insertion, dialysis initiation).

1.1.3 Use Cases and Research Applications

MIMIC-IV is an important data repository for machine learning and health services research. Clinical prediction tasks, such as ICU mortality forecasting, length-of-stay estimation, and readmission risk assessment take advantage of the timestamped vital signs, laboratory values, and treatment interventions [3]. Phenotyping studies get use of diagnosis codes, laboratory trajectories, and clinical notes to identify cohorts with specific physiological conditions. Process mining approaches reconstruct care pathways by sequencing event logs and building knowledge graphs [4]. Moreover, the data set is a good benchmark for algorithms in temporal pattern recognition and decision support based in reinforcement learning.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

1.1.4 Ethics, De-identification, and Access

The MIMIC-IV database is fully de-identified in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Safe Harbor provision [2]. All protected health information has been removed and dates are randomly shifted by a consistent offset per patient. Thus, it does not affect temporal relationships between events, while it prevents reconstruction of actual dates. Direct identifiers such as names and addresses are excluded to ensure patient privacy.

In order to maintain ethical standards for researchers using the MIMIC-IV data, every person accessing the dataset is required to complete the Collaborative Institutional Training Initiative (CITI Program) course titled *Data or Specimens Only Research* and successfully pass associated quizzes. Furthermore, access to the database requires signing the PhysioNet Data Use Agreement (DUA) [5].

For the purposes of this thesis, all necessary steps, including successful completion of the CITI training and agreement to the terms of the DUA were fulfilled. Formal access to the MIMIC-IV database has been requested and granted through PhysioNet. All data handling and analysis procedures that are part of this work strictly adhere to the ethical standards of using this medical data.

1.1.5 Limitations

Diagnoses of patients in MIMIC-IV are not time stamped, which means that the exact time of a diagnosis must be inferred by combining ICD codes with blood culture results and antibiotic order timestamps instead of a standard physiological definition [6]. For example, the onset time of sepsis is often inferred from the available information provided for the patient. For example, by the first hour in which the SOFA score increases by ≥ 2 points, or by the timestamp of an antibiotic order. Therefore, any model trained to predict sepsis is prone to a biased learning of administrative actions instead of the physiological onset. Additionally, labs and vitals are ordered as needed, leading to sparse and irregular ICU records leaving many gaps in a patient's laboratory image. This patchiness requires careful handling of missing values and resampling to fixed intervals.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

1.2 Cohort Selection & Inclusion Criteria

Based on the available MIMIC-IV data, a data cohort was selected. First, all ICU stays with implausible timestamps (for example, negative length of stay or `intime` $\geq$ `outtime`) were excluded. To minimize bias, only each patient's first sepsis related ICU admission was selected. This avoids patients that have chronic sepsis and patients that had multiple transfers between the general ward and the ICU. The following sections outline the criteria for selecting patients and laboratory values in detail, as well as the resulting data distribution.

1.2.1 Identification of Sepsis Cases in MIMIC-IV

Sepsis cases were identified in the `diagnoses_icd` table by their International Classification of Diseases codes (ICD-9-CM and ICD-10-CM). The raw codes were grouped with the Agency for Healthcare Research and Quality's HCUP Clinical Classifications Software (CCS) for ICD-9-CM [7] and its successor, Clinical Classifications Software Refined (CCSR) for ICD-10-CM [8]. Filtering the CCS/CCSR output for the categories *Septicemia* and *Septicemia (except in labor)* yielded 79 distinct ICD codes (Table 1.2).

ICD-9-CM	**ICD-10-CM**
003.1 Salmonella septicemia	A02.1 Salmonella sepsis
020.2 Septicemic plague	A20.7 Septicemic plague
022.3 Anthrax septicemia	A22.7 Anthrax sepsis
036.2 Meningococcemia	A39.2 Acute meningococcemia
	A39.3 Chronic meningococcemia
	A39.4 Meningococcemia, unspecified
038.0 Streptococcal septicemia	A40.0 Sepsis due to *Streptococcus*, group A
	A40.1 Sepsis due to *Streptococcus*, group B
...	
449 Septic arterial embolism	
771.81 Septicemia [sepsis] of newborn	P36.0–P36.9 Sepsis of newborn (various)
	(group B, *Staphylococci*, *E. coli*, etc.)
790.7 Bacteremia NOS	
995.91 SIRS due to infection w/o organ dysfunction	R65.20 Severe sepsis without septic shock
995.92 SIRS due to infection with organ dysfunction	R65.21 Severe sepsis with septic shock

See Appendix 7 for the complete ICD-9-CM and ICD-10-CM mapping.

Table 1.2: CCS septicemia → ICD-9/10 code mapping – first five and last five rows (downloaded July 15, 2024)

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Selecting all hospital admissions containing at least one of these codes produced 22,316 unique entries with a sepsis diagnosis. When these admissions were joined to `icustays`, 17,131 linked ICU stays were obtained (Algorithm 1.1). To ensure that each observation represented an independent clinical episode, stays associated with repeat admissions for the same sepsis event were removed. ICU stays that did not contain laboratory values were deemed insufficient and not included as well due to a lack of a meaningful information.

After these exclusions, the final sepsis cohort comprised 10,711 ICU stays. A second cohort of 10,000 non-sepsis ICU stays was selected with identical selection criteria to preserve consistency. Choosing non-sepsis patients is important to ensure dataset diversity and enable the model to accurately distinguish between sepsis and non-sepsis cases. Although ICD based definitions can miss true sepsis episodes, validation studies report positive predictive values of 80–90% for septicemia codes compared with chart review [9].

Algorithm 1.1: Select sepsis and non-sepsis ICU stays

Input: Table _diagnoses_icd_sepsis (sepsis admissions), table icustays
Output: SepsisICU (10 711 stays), ControlICU (10 000 stays)

```
SELECT DISTINCT
    d.subject_id,
    d.hadm_id,
    i.stay_id,
    i.intime,
    i.outtime,
    i.los
FROM _diagnoses_icd_sepsis d
JOIN icustays i ON d.hadm_id = i.hadm_id
WHERE i.los >= INTERVAL '3 hours' – keep each patients first stay by intime
SELECT DISTINCT
    i.subject_id,
    i.hadm_id,
    i.stay_id,
    i.intime,
    i.outtime,
    i.los
FROM icustays i
LEFT JOIN _diagnoses_icd_sepsis d ON i.hadm_id = d.hadm_id
WHERE d.hadm_id IS NULL AND i.los >= INTERVAL '3 hours'
```

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

1.2.2 Extraction of SOFA-relevant Laboratory Test Values

First, each of the laboratory measurement contained in the SOFA score are identified by their specific `Item ID` codes listed in Table 1.3. These are PaO_2, FiO_2, latelet count, total bilirubin, mean arterial pressure, Norepinephrine rate, Glasgow Coma Scale (GCS), and serum creatinine. These codes map to entries in the MIMIC-IV tables `labevents`, `chartevents`, and `inputevents`. The individual measurements are then restricted to only those values that belong to a previously selected `hadm_id` ICU stays.

The collected measurements are then sorted chronologically according to their timestamps. Each laboratory result is checked for physiological implausibility, such as extremely high or low values that are implausible. Measurements outside the possible range (e.g. platelet counts below 1×10^3 or above $2 \times 10^6\,\mu L^{-1}$, or PaO_2 values exceeding 760 mmHg) were recoded as missing.

Table 1.3: Mapping of SOFA-score components to MIMIC-IV item identifiers (v3.1)

SOFA component	**Variable (unit)**	**Source table**	**Item ID(s)**
Respiratory	PaO_2 (mm Hg)	`labevents`	50821
	FiO_2 (fraction)	`chartevents`	223835
Coagulation	Platelet count ($\times 10^3/\mu L$)	`labevents`	51265
Liver	Total bilirubin (mg/dL)	`labevents`	50885
Cardiovascular	Mean arterial pressure (mm Hg)	`chartevents`	220045
	Norepinephrine rate ($\mu g/kg/min$)	`inputevents`	221906*
CNS	Glasgow Coma Scale (E, V, M)	`chartevents`	223900, 223901, 220739
Renal	Serum creatinine (mg/dL)	`labevents`	50912

*Including additional vasopressor IDs if multiple measurements are used (e.g. epinephrine 221289)

Once validated, each value is assigned an SOFA sub score from 0 to 4 by comparing it against the thresholds in Table 1.4. This mapping is performed for each of the six SOFA values (PaO_2/FiO_2), coagulation (platelet count), liver (bilirubin), cardiovascular mean arterial pressure (MAP) and vasopressor use), neurological (GSC), and renal (creatinine). Since ICU stays often have incomplete lab data (e.g. clinicians might stop ordering additional tests once they have enough information), missing lab measurements were treated explicitly as a separate feature, rather than using imputation methods. Lab values were converted into ordinal SOFA subscores (0–4) to create a reference dataset, which was then used to evaluate how well the DBN predicts changes in those scores.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Component	Range / Condition	Score
Respiratory	$PaO_2/FiO_2 \geq 400$ mm Hg	0
	$PaO_2/FiO_2 < 400$ mm Hg	1
	$PaO_2/FiO_2 < 300$ mm Hg	2
	$PaO_2/FiO_2 < 200$ mm Hg and mechanically ventilated	3
	$PaO_2/FiO_2 < 100$ mm Hg and mechanically ventilated	4
Coagulation	Platelets $\geq 150 \times 10^3/\mu L$	0
	Platelets $100–149 \times 10^3/\mu L$	1
	Platelets $50–99 \times 10^3/\mu L$	2
	Platelets $20–49 \times 10^3/\mu L$	3
	Platelets $< 20 \times 10^3/\mu L$	4
Liver	Bilirubin < 1.2 mg/dL	0
	Bilirubin 1.2–1.9 mg/dL	1
	Bilirubin 2.0–5.9 mg/dL	2
	Bilirubin 6.0–11.9 mg/dL	3
	Bilirubin ≥ 12.0 mg/dL	4
Cardiovascular	MAP ≥ 70 mm Hg	0
	MAP < 70 mm Hg	1
	Dopamine $\leq 5\,\mu$g/kg/min or any-dose dobutamine	2
	Dopamine $> 5\,\mu$g/kg/min or epinephrine $\leq 0.1\,\mu$g/kg/min or norepinephrine $\leq 0.1\,\mu$g/kg/min	3
	Dopamine $> 15\,\mu$g/kg/min or epinephrine $> 0.1\,\mu$g/kg/min or norepinephrine $> 0.1\,\mu$g/kg/min	4
CNS (GCS)	GCS = 15	0
	GCS 13–14	1
	GCS 10–12	2
	GCS 6–9	3
	GCS < 6	4
Renal	Creatinine < 1.2 mg/dL	0
	Creatinine 1.2–1.9 mg/dL	1
	Creatinine 2.0–3.4 mg/dL	2
	Creatinine 3.5–4.9 mg/dL or UO < 500 mL/day	3
	Creatinine ≥ 5.0 mg/dL or UO < 200 mL/day	4

Table 1.4: SOFA sub-score thresholds for laboratory components [10]

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

1.3 Descriptive Analysis of the Study Cohort

1.3.1 Data Preprocessing and Temporal Coverage

All biologically impossible readings were removed. The filtering of implausible values affected less than 0.01% of all rows, and were treated as missing. The continuous predictors were discretized by binning them into equal frequency (quantile) bins. Specifically, for every laboratory variable x the empirical distribution in the cohort was partitioned into $k = 3$ intervals of identical sample size (tertiles) using scikit-learn's `KBinsDiscretizer`. Samples falling into the lowest, middle and highest tertiles were encoded with integer labels $1, 2$, and 3, respectively, while missing measurements were mapped to the dedicated category 0. This quantile based ordinal encoding preserves the natural ordering of the original variables and reduces their dynamic range.

Most ICU stays include multiple laboratory measurements recorded. Table 1.5 shows the number of timestamped lab events recorded during the patients stays. Each timestamp represents the measurement of a relevant SOFA value. Analyzing the distribution of those available timestamped lab events, it is evident that most ICU stays have a substantial amount registered in the MIMIC dataset. The median is six per stay, providing adequate temporal resolution to model sepsis progression.

Timestamps before SOFA	Number of Stays
1	8
2	96
3	369
4	3,452
5	5,313
6	7,629
7	3,741
8	100
9	3

Table 1.5: Number of timestamps per ICU stay

1.3.2 Distribution of Predictors

Among the 20,711 ICU stays analyzed, FiO_2 recordings were available in 12,098 cases and PaO_2 in 15,490. To standardize FiO_2, numeric entries ≤ 1.0 were multiplied by 100% to convert fractions to percentages. Measurements with value 0 were replaced by the ambient 21%, and values exceeding 100% were discarded as clinically implausible. PaO_2 readings were retained only if they fell within the physiologic range of 1–500 mmHg, ensuring biologically realistic input. Figure 1.1 shows the resulting PaO_2/FiO_2 ratio, with the distribution heavily skewed toward 200–600 mmHg. The bin figure also shows that besides having many missing value, the discretized values are equal in frequency.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

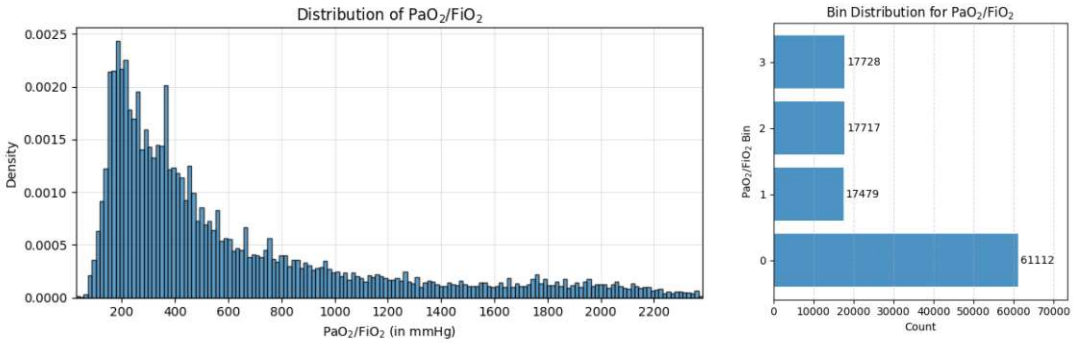


Figure 1.1: Distribution of the PaO_2/FiO_2 values

Total bilirubin, indicative of hepatic function and potential organ dysfunction, was available for analysis in 15,899 stays. Reviewing the distribution of bilirubin (see Figure 1.2) show a right skewed distribution, leaning in the expected range of 0–2 mg/dL. Valus less than 0.1 mg/dL or greater than 30 mg/dL were categorized as erroneous and treated as missing. Notably, only 87 bilirubin measurements were affected by this criterion, underscoring the overall robustness and accuracy of the laboratory data.

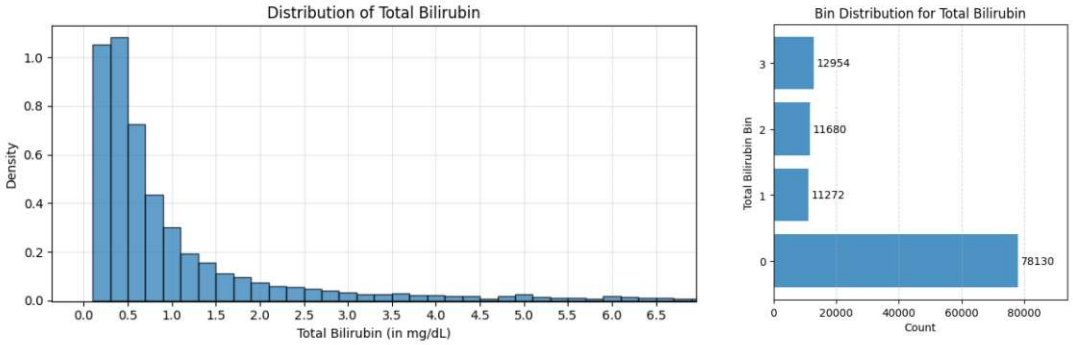


Figure 1.2: Distribution of total Bilirubin measurements

Creatinine measurements, reflective of renal function status, were kept only if they fell between 0.2 and 20 mg/dL. Although elevated creatinine values can occur in critically ill patients, those beyond the established upper threshold of 20 mg/dL were considered biologically implausible outliers. Only 11 measurements exceeded these predefined limits and were recoded as missing. The resulting distribution of creatinine values (see Figure 1.3) exhibit typical normal distribution characteristics, centered on the clinically expected median value of approximately 1 mg/dL.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

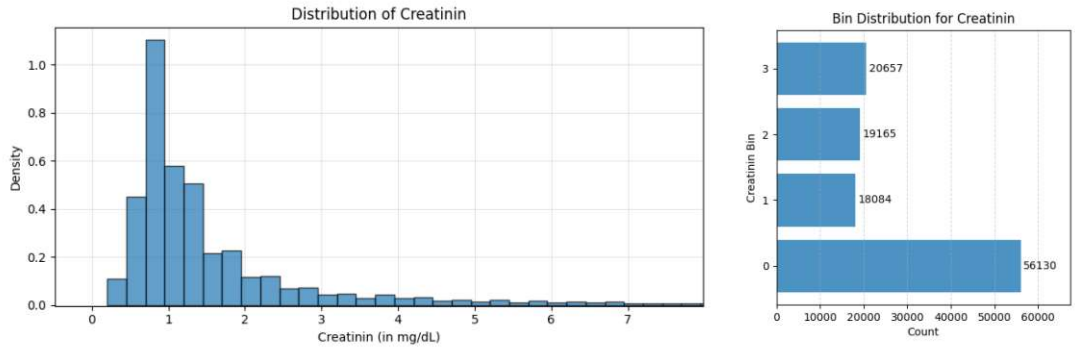


Figure 1.3: Distribution of Creatinin measurements

Mean arterial pressure (MAP), an essential cardiovascular indicator, was nearly universally available across the study cohort, recorded in all 20,711 stays. MAP values underwent plausibility checks, with acceptable physiological bounds set at 30 to 200 mmHg. Only 13 observations fell outside these clinical thresholds and were subsequently treated as missing data. As depicted in Figure 1.4, MAP demonstrated a normal distribution with a peak around 80 mm Hg, consistent with typical ICU targets.

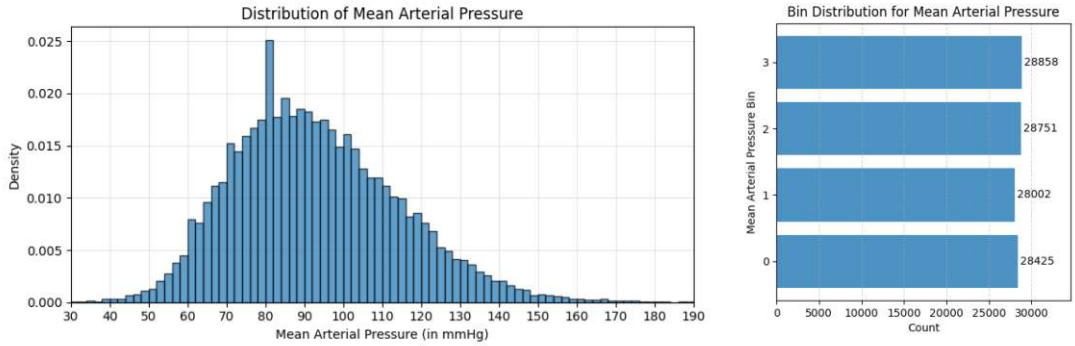


Figure 1.4: Distribution of mean arterial pressure measurements

Platelet counts, crucial for assessing coagulation status and hematologic function, were recorded in 20,393 stays. After imposing clinically justified boundaries of 10 to 1,000 thousand platelets per microliter (thousand/μL), 43 outlier observations were identified and recoded as missing. Figure 1.5 illustrates the platelet count distribution, showcasing a well-defined normal curve with a distinct peak at the typical normal range around 180 thousand/μL.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

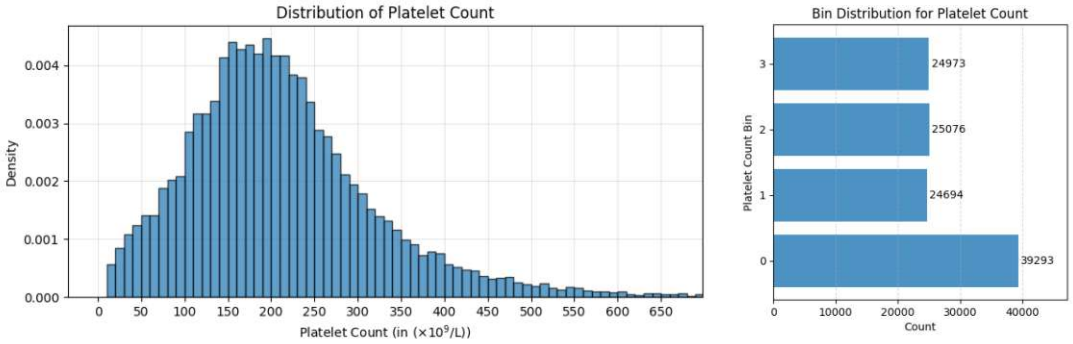
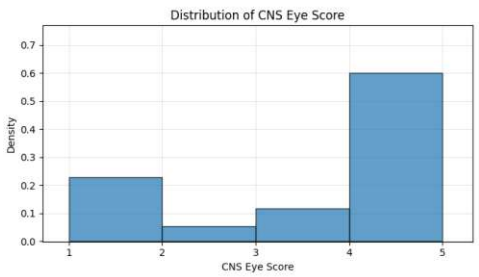
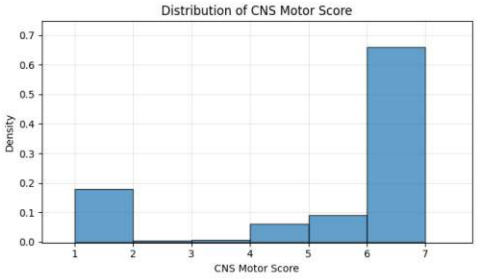


Figure 1.5: Distribution of platelet counts measurements

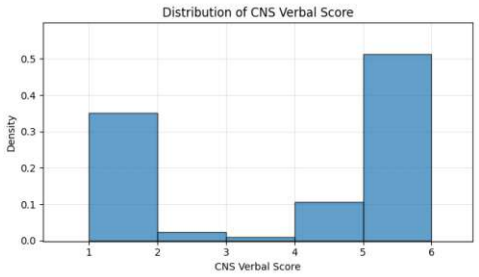
Neurological status was summarized with the Central Nervous System Score(CNS), derived from the three GCS subscores, eye, verbal, and motor. This score consists of three distinct GCS components, namely the eye opening, verbal, and motor responses with unified numeric indicators of neurologic impairment [10]. Missing data in any GCS sub-component was imputed with the normal maximal scores (eye = 4, verbal = 5, motor = 6). Each independent score of the three components is then summed to create the CNS score. Figures 1.6a, 1.6b, 1.6c provide a summary of each GCS sub-component distribution, while Figure 1.6d shows the cumulative CNS score distribution.



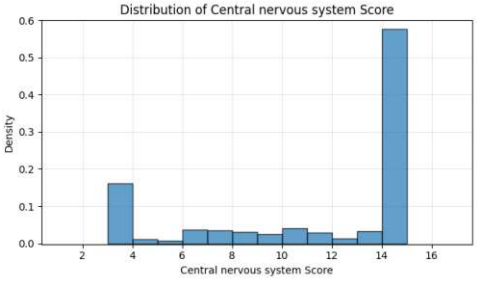
(a) Eye Component



(b) Motor Component



(c) Verbal Component



(d) Combined CNS score

Figure 1.6: Distribution of GCS components and combined CNS score

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

CHAPTER 2

Dynamic Bayesian Network Architecture and Structure Learning

2.0.1 Dynamic Bayesian Networks

A Dynamic Bayesian Network (DBN) is a probabilistic graphical model that represents how a vector of random variables $\mathbf{X}_t = \{X_{1,t}, \ldots, X_{d,t}\}$ changes over discrete time steps $t = 1, 2, \ldots, T$ [11]. First, the process is assumed to be first order Markov, meaning that the distribution of the next state depends only on the current state and is conditionally independent of all earlier states [12], i.e.,

$$P(\mathbf{X}_{t+1} \mid \mathbf{X}_{1:t}) = P(\mathbf{X}_{t+1} \mid \mathbf{X}_t), \tag{2.1}$$

Second, stationarity is assumed, meaning that the way variables influence each other from one time step to the next doesn't change over time. This property implies that

$$P(\mathbf{X}_{t+1} \mid \mathbf{X}_t) = P(\mathbf{X}_{s+1} \mid \mathbf{X}_s) \quad \forall\, s, t \in \{1, \ldots, T-1\}, \tag{2.2}$$

so the transition rules between time steps are fixed and identical for all t, making the reuse of a single set of conditional probability tables (CPTs) across all transitions from timestep t to $t+1$ possible [11]. These assumptions give a compact factorization

$$P(\mathbf{X}_{1:T}) = P(\mathbf{X}_1) \prod_{t=1}^{T-1} P(\mathbf{X}_{t+1} \mid \mathbf{X}_t) \tag{2.3}$$

of the joint distribution.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

The initial distribution $P(\mathbf{X}_1)$ is itself an ordinary Bayesian Network with graph G_0, so that

$$P(\mathbf{X}_1) = \prod_{i=1}^{d} P(X_{i,1} \mid \mathrm{Pa}_{i,1}^{G_0}) \tag{2.4}$$

where $\mathrm{Pa}_{i,1}^{G_0}$ denotes the parents of $X_{i,1}$ in G_0. A network with two time slices $G_{\rightarrow}$ is used to model the change over time. Thus, for each $X_{i,t+1}$ only its parents $\mathrm{Pa}_{i,t+1}^{G_{\rightarrow}}$ in $\mathbf{X}_t$ determine its distribution, so that

$$P(\mathbf{X}_{t+1} \mid \mathbf{X}_t) = \prod_{i=1}^{d} P(X_{i,t+1} \mid \mathrm{Pa}_{i,t+1}^{G_{\rightarrow}}) \tag{2.5}$$

holds.

Because of stationarity, the directed graph structure $G_{\rightarrow}$ and its CPTs remain identical for every timestep. This means that instead of estimating tables at each slice, one fixed set of parent–child relationships and probability parameters for all transitions from t to $t+1$ are used. This approach is also called *parameter tying* [11] and it reduces the total number of parameters that need to be estimated.

Inference in a DBN primarily uses the filtering distribution which yields the posterior of the hidden state at time t given all observations up to that point [13]. Although a smoothing distribution can incorporate future data, it is not required here because the goal is continuous sepsis prediction on incoming lab values without insight into future lab values. Therefore the focus of the inference will be exclusively on computing

$$P(\mathbf{X}_{t+1} \mid \mathbf{X}_{t:t+1}^{\mathrm{obs}}), \tag{2.6}$$

treating the binary sepsis indicator at timestep t (S_t) as an unobserved variable whose probability is updated as new lab values arrive.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

2.1 Time Slice Construction of ICU Data

Laboratory measurements usually arrive at irregular, patient specific timestamps. Therefore each admission is unfolded into consecutive observation pairs $(t-1, t)$. Let S_t be the binary sepsis indicator and $\mathbf{L}_t = \{L_{1,t}, \ldots, L_{n,t}\}$ the n laboratory values at time t. After transformation, every training row contains

$$S_{t-1},\, L_{1,t-1}, \ldots, L_{n,t-1},\; S_t,\, L_{1,t}, \ldots, L_{n,t}, \tag{2.7}$$

meaning that every data entry contains laboratory values from the previous timestep as well as the current measurement. This is done by sorting all data entires chronologically within each `hadm_id`. Afterwards, shifting the data by one timestep makes a copy of every column, moved down by one row, so the value at $t-1$ sits next to the value at t. First, admissions with a single record are discarded, because no $(t-1, t)$ pair can be formed. The temporal data table can go straight into a hill-climbing learner and still keep it ordered, namely numbers from slice $t-1$ are used to predict slice t.

2.2 Structure Learning in Dynamic Bayesian Networks

2.2.1 Hill-Climbing Structure Learning

The dependency structure of the DBN is not known a priori and must be learned from the available data. An exhaustive search for an appropriate Directed Acyclic Graphs (DAGs) grows exponentially with the number of variables. Thus an approximation of an optimal structure must be done. For this a greedy hill-climbing heuristic was used. The algorithm is guided by a scoring function that balances data fit and model complexity, while the search space is constrained by physiologically plausible edge constraints [14] (see Section 2.2.2).

The Bayesian Information Criterion for discrete data (*BIC-d*) scores a DAG G by balancing data fit against model complexity [15]. For each variable X_i, let r_i be its number of states and let its parent set Pa_i^G have joint cardinality

$$q_i \;=\; \prod_{X_j \in \mathrm{Pa}_i^G} r_j \tag{2.8}$$

where N_{ijk} is the number of two-slice samples in which $X_i = k$ and its parents are in configuration j. The maximum-likelihood estimate of the conditional probability parameters is given by

$$\hat{\theta}_{ijk} \;:=\; \frac{N_{ijk}}{\sum_{k'=1}^{r_i} N_{ijk'}}. \tag{2.9}$$

With N being the total number of two-slice observations, the local log-likelihood factor is factorized as

$$\log P(X_i \mid \mathrm{Pa}_i^G, \hat{\theta}) \;=\; \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} N_{ijk} \log \hat{\theta}_{ijk}, \tag{2.10}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

and the global score is computed by summing the local contributions and subtracting the penalty for parameters $|\Theta_i| = (r_i - 1)q_i$, so that

$$\text{BIC-d}(G) \;=\; \sum_{i=1}^{d} \Big[\log P(X_i \mid \text{Pa}_i^G, \hat{\theta}) \;-\; \tfrac{1}{2} \, |\Theta_i| \, \log N \Big]. \tag{2.11}$$

Because the score decomposes by node, updating a single edge takes $O(q_i r_i)$ time for the affected nodes. Consequently, each evaluation is fast and the overall search runs in time roughly linear in the number of moves.

2.2.2 Domain-Guided Edge Constraints and Structure Search Method

The hill-climbing algorithm receives a blacklist of edges containing edges that are physiologically impossible. Within a single time t slice the sepsis node S_t can influence every laboratory value $L_{i,t}$, but no lab is allowed to point back to sepsis, leading to

$$P(S_t, \mathbf{L}_t) = P(S_t) \, P(\mathbf{L}_t \mid S_t), \tag{2.12}$$

a factorization of the joint distribution:

This makes the labs conditionally independent once sepsis status is fixed. Another constraint is that the causality only moves forward. Edges can run from variables at time t to variables at time $t+1$, but never the reverse, preserving the first order Markov form

$$P(X_{1:T}) = P(X_1) \prod_{t=2}^{T} P(X_t \mid X_{t-1}), \qquad X_t = \{S_t, \mathbf{L}_t\},$$

where each state X_t consists of sepsis variable S_t and the laboratory measurements $\mathbf{L}_t$.

With this blacklist in place, structure learning starts from an empty graph. The greedy hill-climbing suggests single edge additions, deletions, or reversals and accepts a move only when it raises the BIC-d score by at least $\varepsilon = 10^{-4}$. In order for the search to not get stuck in local optima, the edge of each accepted move is written to a *tabu list*. While an edge is on that list any move that would lead to a creation of a graph similar to its previous form is deemed taboo and skipped, even if it offers a modest score gain [16]. This usually affects the exact inverse of the last changed edge, preventing the search from oscillating between nearly identical graphs and forces it to explore new neighborhoods. Because every candidate move is still checked against the physiological blacklist, all intermediate graphs remain plausible.

Algorithm 2.1 evaluates at most $3d(d-1)$ moves per iteration. The three elementary operations (add, delete, reverse) are applied to every ordered pair of vertices in a graph with d nodes. First, any move that would create a directed cycle is rejected. Second, the blacklist eliminates edges that contradict known physiology. Third, recently edited edges are removed from the taboo queue, so the search cannot undo its last steps. If changes pass those restrictions, they are scored with BIC-d, and the best one is accepted if it improves the score by more than $\varepsilon = 10^{-4}$.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Algorithm 2.1: GreedyHillClimb($G^{(0)}, \mathcal{C}, \varepsilon$)

Input: Initial DAG $G^{(0)}$; whitelisted edge set $\mathcal{C}$; BIC-d gain threshold $\varepsilon > 10^{-4}$
Output: Locally optimal DAG $\widehat{G}$

$\ell \leftarrow$ BIC-d$(G^{(0)})$
$k \leftarrow 0$
repeat
  // $\Delta_{\max} \leq \varepsilon$
  $\Delta_{\max} \leftarrow 0$; $G^{\star} \leftarrow G^{(k)}$
  foreach *operation* $m \in \{\text{add}, \text{del}, \text{rev}\}$ **do**
    foreach *vertex pair* (u, v) *allowed by* $\mathcal{C}$ **do**
      $G' \leftarrow$ apply m on (u, v) in $G^{(k)}$
      if G' *is acyclic* **then**
        $\Delta \leftarrow$ BIC-d$(G') - \ell$
        if $\Delta > \Delta_{\max}$ **then**
          $\Delta_{\max} \leftarrow \Delta$
          $G^{\star} \leftarrow G'$
          $(u^*, v^*, m^*) \leftarrow (u, v, m)$
        end
      end
    end
  end
  if $\Delta_{\max} > \varepsilon$ **then**
    $G^{(k+1)} \leftarrow G^{\star}$; $\ell \leftarrow \ell + \Delta_{\max}$; $k \leftarrow k + 1$
  end
until no admissible move improves the score
return $\widehat{G} = G^{(k)}$

2.2.3 Stability Selection and Pruning

To avoid overfitting, the learning phase is combined with a bootstrap approach to stability selection [17]. Ten bootstrap replicas of the training set are created, a DAG is learned on each replica, and all edges are recorded. Only edges that occur in at least 80% of replicas are kept as stable edges.

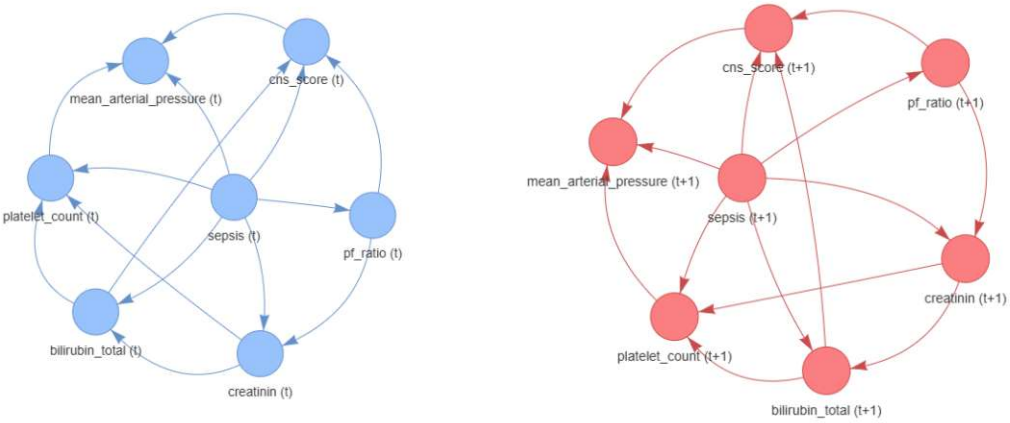
The resulting graph then undergoes backward pruning, computing any edge whose removal worsens the BIC-d score by less than $\Delta_{\text{BIC}} = -100$. Those edges that have only a minimal impact on the BIC-d score are consequently removed in order to have a confident but not overly complex graph structure.

The finished model is therefore (i) constrained by known physiology, (ii) discovered through a tabu-guided BIC search, (iii) stabilized via bootstrapping, and (iv) trimmed by BIC pruning, making it coherent, clinically viable network ready for sepsis prediction.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

2.3 Final DBN Structure

Intra-slice structure Figures 2.1a shows the causal dependencies within the first time point t. Similarly, Figure 2.1b depicts the dependencies of time slice t+1. Consistent with the medical constraints introduced earlier, sepsis status causally influences laboratory measurements recorded at the same time. This pattern is the same for both slice t and slice t+1.



(a) Intra-slice network at time t (b) Intra-slice Network at Time t+1

Figure 2.1: Causal dependencies between sepsis and lab variables at (a) t and (b) t+1.

Inter-slice structure Figure 2.2 depicts the temporal dependencies between both time slice graphs t and $t+1$. These dependencies respect the first order Markov assumption - variables at time t influence their own as well as states at time $t+1$. These links describe how clinical measurements evolve over time and supply the information required for sequential inference. The lone edge linking sepsis at t to sepsis at $t+1$ shows that this diagnostic variable is the prediction target and hence should not form connections across timeslices. Instead, it only drives the laboratory measurements within each individual slice.

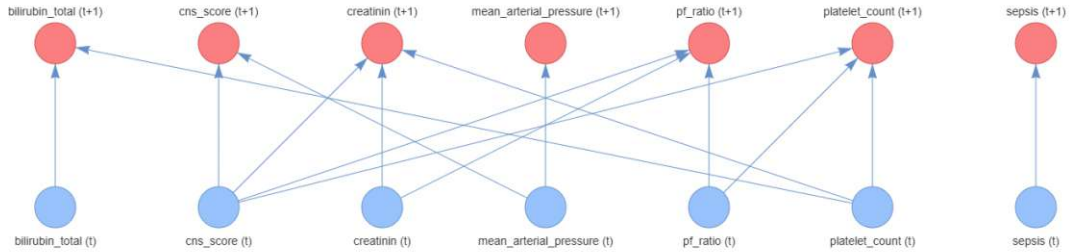


Figure 2.2: Inter-slice network showing directed temporal dependencies from slice t to slice t+1

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Full network structure Overlaying both the intra- and inter-slice edges, results in a complete depiction of the DBN and giving a single view in Figure 2.3. The diagram illustrates all temporal causal pathways, providing a comprehensive graphical summary suitable for clinical interpretation and predictive inference.

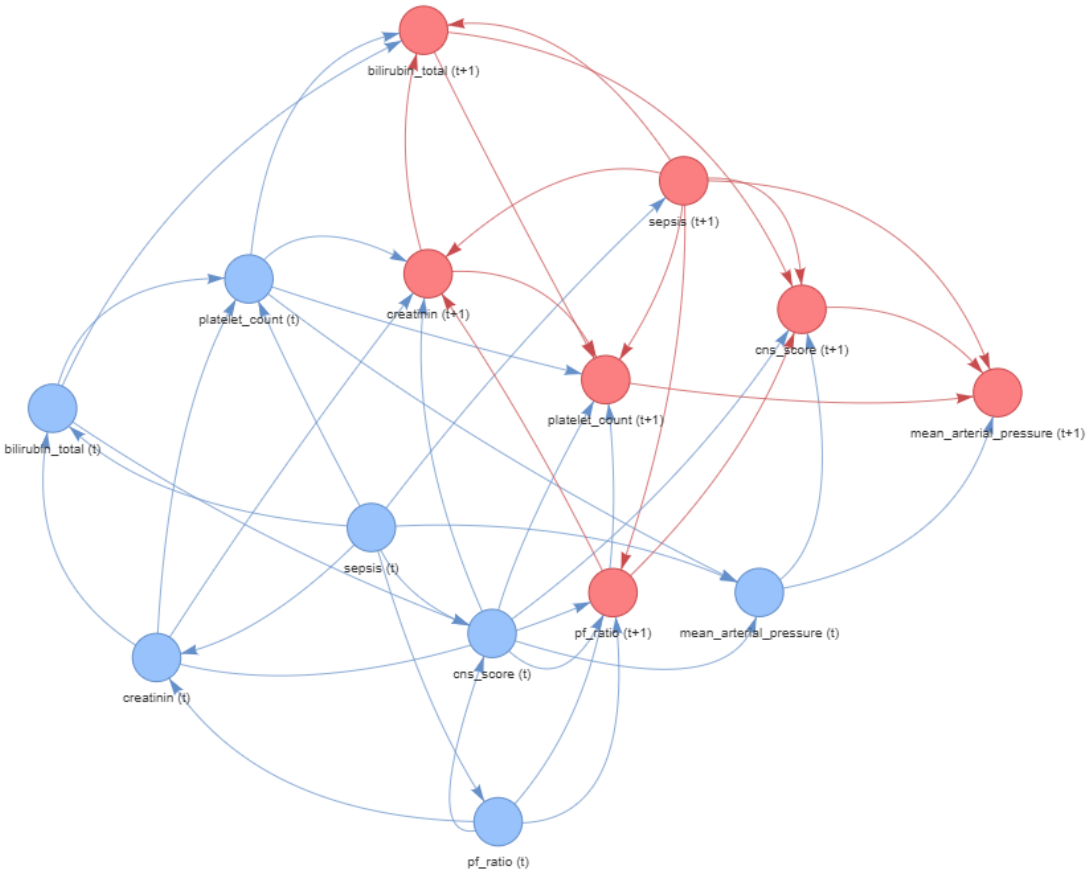


Figure 2.3: Complete DBN integrating intra-slice and inter-slice edges; the graph visualises the full set of temporal and contemporaneous relationships

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

CHAPTER 3

Dynamic Bayesian Network Parameter Estimation

After determining the optimal structure of the DBN in Chapter 2, the next step is parameter estimation to compute the probability relationships represented by the network's edges. These relationships are expressed as Conditional Probability Distributions (CPDs). This chapter explains the methodological details, and practical application of parameter estimation using Maximum Likelihood Estimation (MLE).

3.1 Maximum Likelihood Estimation (MLE)

Maximum Likelihood Estimation (MLE) is a statistical approach used for estimating parameters of probability distributions by maximizing the likelihood function $L(\theta)$. Given a dataset $D = \{X^{(1)}, X^{(2)}, \ldots, X^{(N)}\}$, consisting of N independent and identically distributed samples, the likelihood function $L(\theta)$ is defined as

$$L(\theta) = P(D|\theta) = \prod_{m=1}^{N} P(X^{(m)}|\theta), \tag{3.1}$$

and MLE aims to find parameters $\hat{\theta}$ that maximize the likelihood above, by computing

$$\hat{\theta} = \arg\max_{\theta} L(\theta). \tag{3.2}$$

In practice, the log-likelihood function is typically maximized for convenience [18], so that

$$\ell(\theta) = \log L(\theta) = \sum_{m=1}^{N} \log P(X^{(m)}|\theta). \tag{3.3}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

For Bayesian networks, each node X_i with parents $Pa(X_i)$ has parameters

$$\hat{\theta}_{ijk} = \frac{N_{ijk}}{\sum_{k'} N_{ijk'}}, \tag{3.4}$$

represented as Conditional Probability Tables (CPTs) [19] and N_{ijk} is the count of occurrences where node X_i is in state k given its parent configuration j.

3.2 Application to the DBN Model

In this work's DBN, the parameters are estimated across both time slices, the initial slice $t = 0$ and subsequent slices $t \geq 1$. Thus, the DBN parameters are classified into three categories [13]:

- **Initial CPDs** $P(X_0)$: Initial distributions without parents.
- **Intra-slice CPDs** $P(X_0 \mid Pa(X_0))$: Dependencies within the initial time slice.
- **Transition CPDs** $P(X_{t+1} \mid Pa(X_t) \cup Pa(X_{t+1}))$: Dependencies on variables from slice t and contemporaneous parents in slice $t + 1$.

MLE was applied to each CPD type separately using observed frequency counts from the flattened dataset. In code the computation was performed by `pgmpy`'s implementation of `MaximumLikelihoodEstimator` [20].

To prevent zero probabilities for rare or unseen parent–child configurations, we apply Laplace smoothing to each row of the Conditional Probability Table [21]. Concretely, a small constant $\alpha = 10^{-6}$ was added to every count before normalization. This guarantees that every possible state retains a non-zero probability, improving numerical stability and making our estimates more robust.

After smoothing, the full set of CPDs was attached to the template network and verified that every table sums to one and that cardinalities are consistent with the learned graph. The log-likelihood

$$\ell(\hat{\theta}) = \sum_{m=1}^{N} \log P(\mathbf{X}^{(m)} \mid \hat{\theta}),$$

of the fitted model was saved so later calculations of the BIC was possible and comparison to different models was enabled. Once all CPDs have been fitted and validated, the parameterized DBN was passed to `DBNInference`, which provides methods for exact filtering, smoothing, and prediction on new patient records. This allows to perform inference tasks with the learned model, and run prediction experiments described in the evaluation chapter.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

3.3 Prediction with DBN

The DBNs inference uses forward inference, allowing the incoming probabilities in $t+1$ to update based on sequentially observed data in t. Specifically, sepsis predictions at time t are calculated as

$$P(\text{sepsis}_t | evidence_{0:t}) = \frac{P(\text{sepsis}_t, evidence_{0:t})}{P(evidence_{0:t})} \tag{3.5}$$

posterior distributions given all observed lab evidence

The inner procedure PredictSepsis that processes one admission at a time can be seen in Algorithm 3.1. At each timestamp t, it appends the new observed lab values to the already existing evidence ev. The DBN's forward inference routine computes the posterior $P(\text{sepsis}_t = 1 \mid \text{ev})$. Since only $|\mathcal{L}|$ new entries are added per step and only the sepsis node was queried, the cost of calculating each patient was $\mathcal{O}(T\,|\mathcal{L}|)$. The outer loop then applies this procedure to every admission in the validation data, storing each patient's probability time series in the mapping Pred.

Algorithm 3.1: DBN based sepsis prediction for a test cohort

Input: Test dataframe *df*, lab set $\mathcal{L}$, DBN inference object Inf
Output: Mapping Pred from admission ID to time-stamped sepsis probabilities

Procedure *PredictSepsis*(*patient*, $\mathcal{L}$, Inf)**:**
  ev $\leftarrow \emptyset$;
  for $t \leftarrow 0$ **to** $T-1$ **do**
    foreach $\ell \in \mathcal{L}$ **do**
      ev$[(\ell, t)] \leftarrow$ *patient*$[\ell, t]$;
    end
    $q \leftarrow$ Inf.forward_inference($\{(\text{sepsis}, t)\}$, ev);
    $p_t \leftarrow q[(\text{sepsis} = 1)]$;
  end
  return $\{p_t\}$;
Pred $\leftarrow \emptyset$;
foreach *admission ID h in df*.index **do**
  patient $\leftarrow$ *df*$[h, \mathcal{L}]$;
  Pred$[h] \leftarrow$ PredictSepsis(*patient*, $\mathcal{L}$, Inf);
end
return Pred;

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

CHAPTER 4

Evaluation of DBN and SOFA Scoring

4.1 Evaluation Dataset

For the evaluation of how capable the DBN is of accurate predictions in comparison to the traditional SOFA scoring method, 30% of the total dataset was reserved as test set, consisting of around 2,700 non-septic ICU stays and 3,508 septic ICU stays. Particular attention is given to early detection capabilities and lead time gains, as well as to accuracy, precision, recall, and F1-Scores and adequately interpreted.

4.1.1 Classification Performance

Classification performance was assessed using a probability threshold of 0.5 for the DBN and a severity threshold of SOFA scores equal to or greater than 2. The confusion matrices for both methods are summarized in Table 4.1 and Table 4.2, providing a comparison of accuracy, precision, recall, and F1-Scores.

Table 4.1: Confusion Matrix – DBN (Threshold = 0.5)

	Predicted Negative	Predicted Positive
True Negative	788	1912
True Positive	290	3218

Table 4.2: Confusion Matrix – SOFA (Threshold ≥ 2)

	Predicted Negative	Predicted Positive
True Negative	876	1824
True Positive	403	3105

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table 4.3 presents the performance metrics computed from these matrices.

Table 4.3: Performance metrics comparison (DBN vs. SOFA)

Metric	DBN	SOFA
Accuracy	0.6454	0.6412
Precision	0.6275	0.6301
Recall	0.9174	0.8850
F1 Score	0.7457	0.7362

Both the DBN and SOFA methods show similar results of accuracy, precision, recall, and F1-Scores, with the DBN having slightly improved recall (0.9174 vs. 0.8850) and F1-Score (0.7457 vs. 0.7362). Looking at the confusion matrix, a similar ratio of false positives, and false negatives appears. However, the DBN produces fewer false negatives (290 versus 402). In a clinical setting, this reduction in missed cases can have a big impact on patient outcomes and care effectiveness. The similarity in predictive performance suggests that the underlying data distribution may limit improvements.

4.1.2 Lead Time Analysis

Since in clinical prediction, early diagnosis is crucial, this chapter provides a detailed analysis of predictive lead time. When comparing the exact prediction timings by the number of timestamped labs needed for the diagnosis between DBN and SOFA methods, the DBN significantly outperformed SOFA in terms of early detection. Specifically, the DBN predicted sepsis earlier than SOFA in 2,078 cases, compared to only 157 cases where SOFA predicted it earlier, with 649 instances predicting it at the same timestamp. Figure 6.1 shows the distribution of lead times. The median lead time in terms of lab tests is approximately 2 tests earlier, and the mean is 1.6 labs, suggesting that the DBN method can significantly accelerate the detection of sepsis.

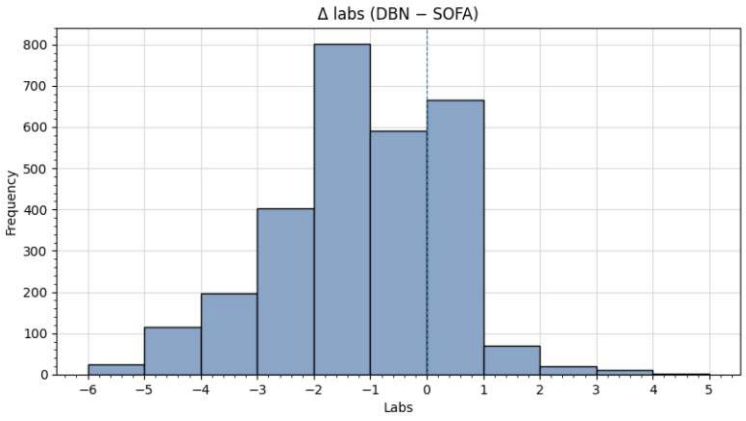


Figure 4.1: Distribution of DBN lead in number of lab tests

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

To approximate lead time in hours, the individual time differences between each timestamped lab value and the diagnosis onset were calculated as well. This measurement of the lead times in hours revealed that the DBN predictions precede SOFA predictions by a median of approximately 0.8 hours, while the mean is 4.2 hours as shown in Figure 6.2. This lead time can be critically important in intensive care settings, potentially enabling earlier clinical interventions.

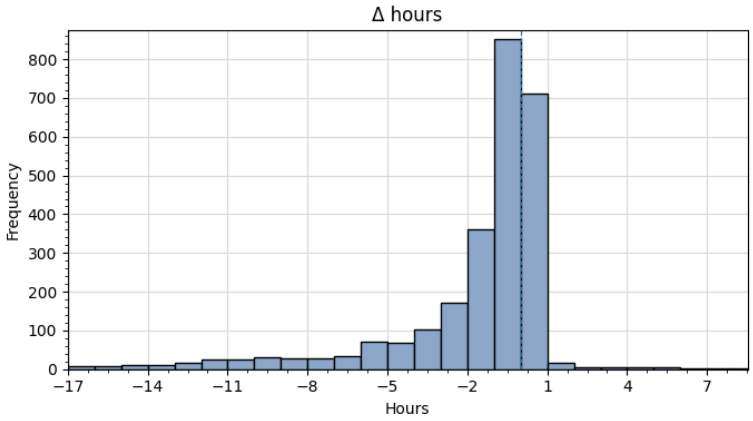


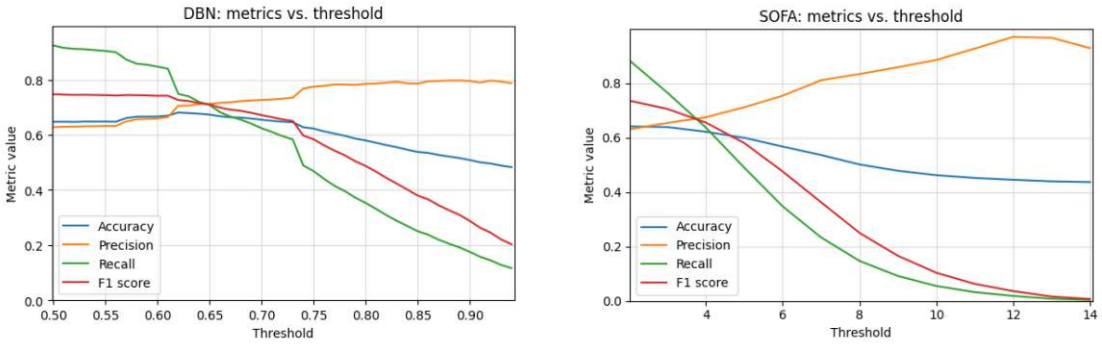
Figure 4.2: Distribution of DBN lead in hours

4.1.3 Impact of Probability Thresholds

Since those scores only apply to predictions with a confidence of at least 0.5, it is difficult to understand how confident the DBN is in its prediction. To better understand the sensitivity of DBN's predictive lead times to different probability thresholds, varying threasholds were applied during the prediction process.

Figure 4.3 shows the change in accuracy, precission, recall and F1-Score with varying threasholds set for both the DBN and SOFA score. Both panels show that model behaviour depends strongly on the chosen threshold. For the DBN, thresholds below 0.60 keep recall close to 0.9, meaning almost every true case is detected, but precision stays around 0.65, so many alarms are false. After that, recall falls faster than precision rises, so both the F1-score and accuracy drop. The SOFA curves display the same pattern. With thresholds near 2 to 4 the score still identifies most true cases while keeping false positives low. At about 7, recall and F1-score decline even though precision continues to improve. Raising the threshold to require more confidence leads to fewer positive calls, so the model misses more true cases and overall accuracy decreases.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.



(a) DBN performance metrics vs. threshold (b) SOFA performance metrics vs. threshold

Figure 4.3: Comparison of DBN and SOFA Sepsis-Specific Metrics Across Thresholds.

Figure 4.4 shows how the proportion of cases where DBN predicts sepsis earlier varies with increasing DBN prediction thresholds, while the SOFA threshold stays constant at 2. As the DBN threshold increases, the frequency of early predictions decreases. At a confidence threshold around 0.71, SOFA begins to outperform DBN in terms of early prediction frequency. This indicates a trade-off between the early prediction advantage of the DBN and its prediction confidence, meaning while DBN detects sepsis earlier, it does so at lower confidence levels.

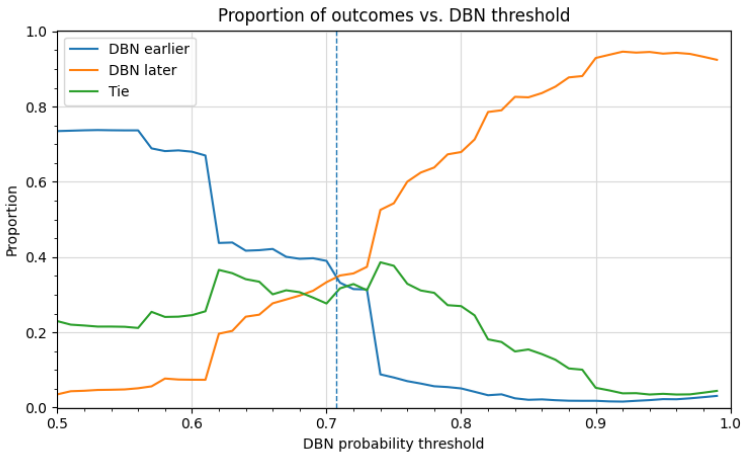


Figure 4.4: Proportion of outcomes relative to varying DBN thresholds

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

The SOFA score itself is adjustable as well and represents different severity of clinical signals. Thus, an extended analysis across varying thresholds of both DBN and SOFA scores was performed, resulting in the heatmap in (see Figure 4.5). It shows how DBN leads in predicting sepsis at lower thresholds, especially for moderate clinical severity indicators of SOFA scores between 2 and 5. However, with stricter diagnostic criteria as both DBN and SOFA thresholds increases the advantage in early prediction diminishes significantly. This highlights the DBN's ability to predict moderate risk diagnosis, while being less confident at early sepsis offset.

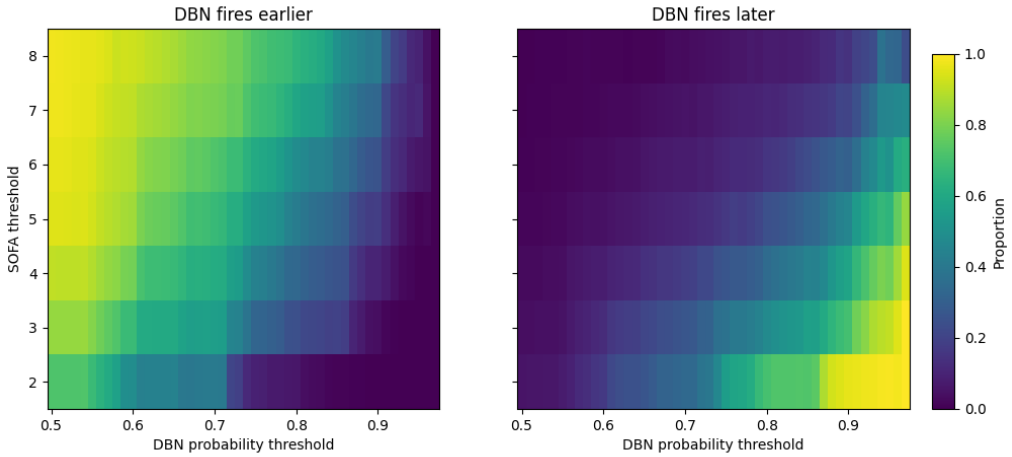


Figure 4.5: Heatmap of proportions where DBN predicts sepsis earlier relative to varying DBN and SOFA thresholds

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

CHAPTER 5

Theory and Methodology of Value of Information (VoI)

To improve early detection of sepsis using a DBN, a Value of Information (VoI) method was applied [22–24]. Specifically, it determines which additional laboratory tests would most effectively improve the predictive confidence of sepsis detection. Specifically, it determines which additional laboratory tests would most effectively improve the predictive confidence of sepsis detection.

VoI quantifies uncertainty by calculating the information gain [25]. Information gain captures the reduction in entropy in the probability distribution of sepsis after observing an additional laboratory measurement. Each sequential measurement selection aims to maximize this reduction in uncertainty.

Mathematically, the posterior probability of sepsis given the current evidence set E was computed using Bayesian inference [13, 26] as

$$P(\text{sepsis} = 1 \mid E) = \text{DBN Inference}(E),$$

where the right side denotes inference in the trained DBN given evidence E. The concept of information gain (IG) was mathematically defined as the reduction in entropy achieved by observing a new piece of evidence E_{new}, namely

$$\text{IG}(E, E_{\text{new}}) = H[P(\text{sepsis} = 1 \mid E)] - H[P(\text{sepsis} = 1 \mid E, E_{\text{new}})]\,,$$

where $H(p)$ represents the binary entropy function [25] and is calculated as

$$H(p) = -p \log_2(p) - (1-p) \log_2(1-p),$$

so that lower posterior uncertainty after adding E_{new} yields higher information gain. To avoid taking $\log_2$ of exactly zero, which would produce undefined or infinite values, each

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

probability p is clipped to

$$p \in [10^{-12}, 1 - 10^{-12}],$$

which guarantees that both $\log_2(p)$ and $\log_2(1-p)$ remain finite while having a negligible effect on the resulting entropy values.

The practical implementation of this theory involves a sequential greedy algorithm [23]. Initially, all candidate laboratory measurements are considered unobserved, represented by placeholder values indicating no measurement. Only those laboratory measurements that were taken for a given patient during the entire stay were considered. Algorithm 5.1 first sets every laboratory node at slice 0 to the value 0, thereby marking all tests as unmeasured. From the patient record it keeps only those variables that contain at least one non-zero entry as the candidate set. A baseline posterior probability of sepsis, p_{base}, was obtained from the DBN under this empty evidence.

At each iteration of the initial loop every hidden candidate was replaced by its earliest real measurement. Forward inference was performed, and the corresponding reduction in binary entropy was recorded. The measurement that yields in the greatest reduction was chosen, removed from the candidate list, and its posterior becomes the new baseline. The loop ends when no remaining test provides positive information gain or when the posterior exceeds the decision threshold τ.

Because the real measurements are revealed in an optimised order rather than in their historical order, the procedure constructs an alternative timeline. Each step stores the selected test, its information gain, and the updated posterior. By comparing the timestamp VoI sepsis posterior first crosses a certain evaluation threshold with when the original lab schedule does, the method shows how much earlier sepsis could be flagged by reordering the tests.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Algorithm 5.1: VoI-based sequential measurement selection for early sepsis prediction

Input: Patient dataframe *patient_df*, lab set $\mathcal{L}$, DBN inference object Inf, prediction threshold τ

Output: Sequential steps with VoI analysis results

Procedure *SimulatePatient*(*patient_df*, $\mathcal{L}$, Inf, τ)**:**

ev $\leftarrow \{(\ell, 0) : 0 \mid \ell \in \mathcal{L}\}$;
results $\leftarrow \emptyset$;
candidate_labs $\leftarrow \{\ell \mid$ first non-zero measurement of ℓ exists$\}$;
$p_{\text{base}} \leftarrow$ Inf.forward_inference($\{(\text{sepsis}, 0)\}$, ev)[sepsis $= 1$];
append initial baseline step to results;
while candidate_labs $\neq \emptyset$ **do**
  best_gain $\leftarrow -\infty$;
  best_lab $\leftarrow$ None;
  foreach $\ell \in$ candidate_labs **do**
    $\text{ev}_{\text{test}} \leftarrow$ ev;
    update $\text{ev}_{\text{test}}[(\ell, 0)]$ with earliest measurement of ℓ;
    $p_{\text{new}} \leftarrow$ Inf.forward_inference($\{(\text{sepsis}, 0)\}$, ev_{test})[sepsis $= 1$];
    IG $\leftarrow H(p_{\text{base}}) - H(p_{\text{new}})$;
    if $IG >$ best_gain **then**
      best_gain $\leftarrow$ IG;
      best_lab $\leftarrow \ell$;
      $p_{\text{best}} \leftarrow p_{\text{new}}$;
    end
  end
  if best_lab $=$ *None* **then**
    break
  end
  update ev with best_lab measurement;
  remove best_lab from candidate_labs;
  append step details to results;
  $p_{\text{base}} \leftarrow p_{\text{best}}$;
end
return results;

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

CHAPTER 6

Evaluation of the DBN Model Based on Value of Information (VoI)

6.0.1 Evaluation Dataset

A 30% hold-out of the full cohort was reserved for testing the VoI-augmented DBN, mirroring the previous evaluation. This test set comprised approximately 2,700 non-septic ICU stays and 3,508 septic ICU stays. Emphasis was placed on comparing the lead time for sepsis prediction between the VoI-enhanced DBN and the SOFA scoring method.

6.0.2 Lead-Time Analysis

When comparing the exact prediction timings by the number of timestamped labs required for diagnosis, the VoI-enhanced DBN again outperformed the SOFA method in terms of early detection. Specifically, sepsis was predicted earlier by the VoI approach in 2,606 cases, compared with 2,078 cases for the baseline DBN, indicating a substantial improvement. Figure 6.1 shows the distribution of these lead times. The median lead time in terms of lab tests remains 2 tests earlier, while the mean lead time increases to 2 labs, further demonstrating that the VoI method can significantly accelerate sepsis detection beyond the original DBN's average of 1.6 labs.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

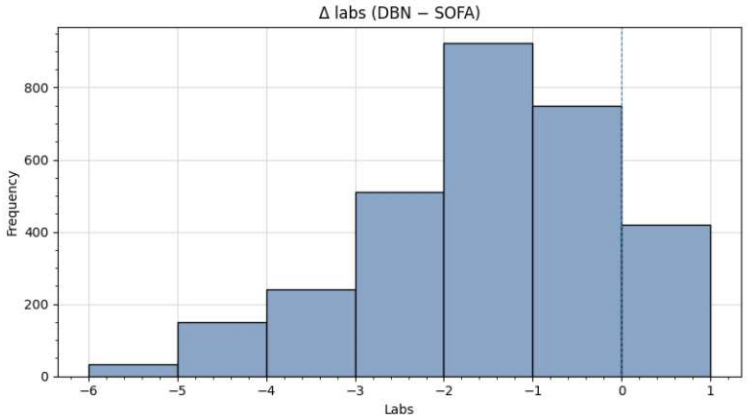


Figure 6.1: Lead-time advantage measured in number of lab results required (DBN-VoI vs. SOFA)

6.0.3 Lead Time in Hours

Lead time in hours was estimated to further quantify the early detection advantage. Because the constructed timelines did not correspond to real clock times, the average interval between consecutive lab tests was used to convert the lab-based lead times into hours. Using this method, VoI-enhanced DBN predictions were found to precede SOFA predictions by a median of 1.5 hours and a mean of 7.1 hours (see Figure 6.2). In comparison, the standard DBN achieved a median lead time of approximately 0.8 hours and a mean of 4.2 hours, indicating that the VoI approach yielded an almost double the average early warning window. Such an increase in lead time could allow for significantly earlier clinical intervention, potentially improving patient outcomes and reducing the risk of sepsis progression.

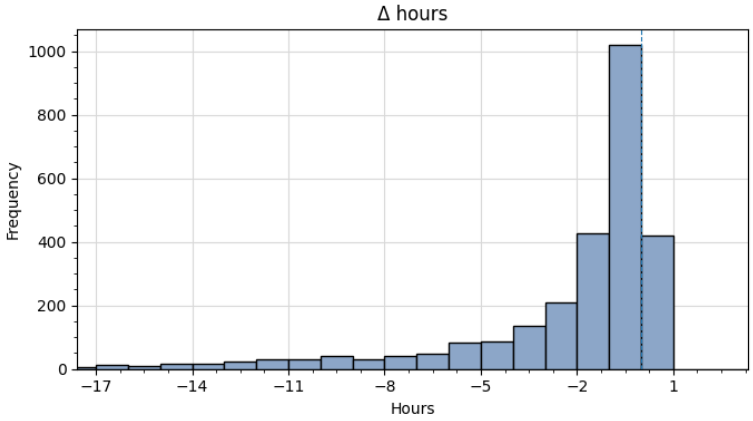


Figure 6.2: Lead-time advantage measured in hours (DBN-VoI vs. SOFA)

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

6.0.4 Prediction Curves

Figures 6.3 and 6.4 illustrate the patient sepsis probability over time for the SOFA method and the VoI-enhanced DBN. It can be observed that the VoI approach reaches diagnostic confidence thresholds significantly earlier than SOFA, showing a rapid increase in confidence after a few lab values. In contrast, the SOFA score rises steadily with each additional lab. This accelerated confidence gain results from selecting the most informative labs, enabling the earliest possible diagnosis.

Figure 6.3 highlights the importance of selecting the most informative laboratory values. The first chosen lab produced the largest increase in diagnostic confidence; if other labs had been selected first, confidence would have stagnated and the diagnostic threshold might have been reached too late. In the SOFA method, three lab values are required to exceed a score of 2, resulting in delayed diagnosis. The VoI approach therefore significantly reduces time to diagnosis and the number of lab tests needed.

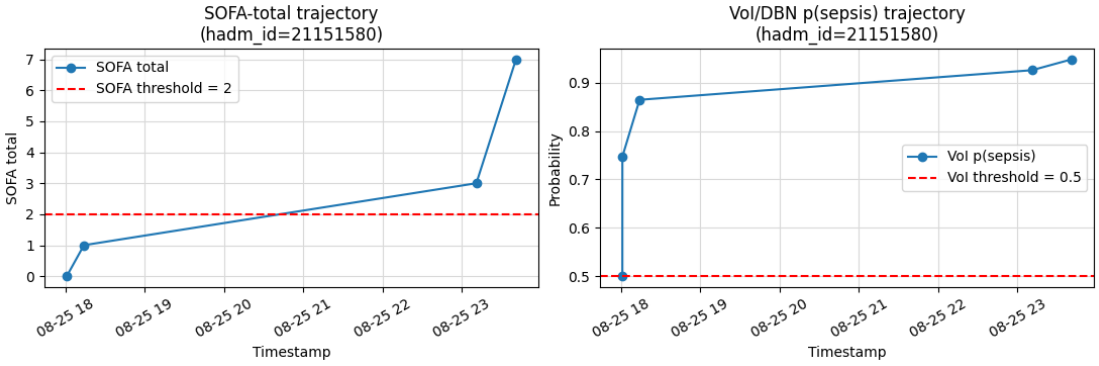


Figure 6.3: Trajectory comparison for patient `hadm_id`=21151580

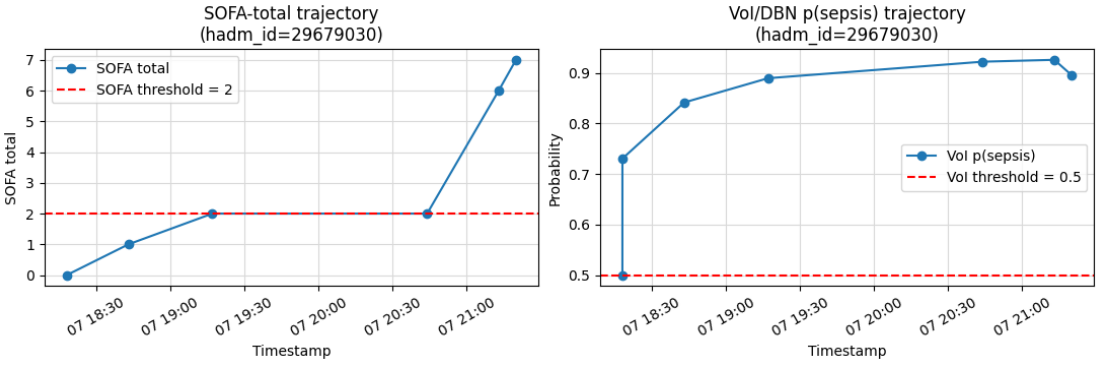


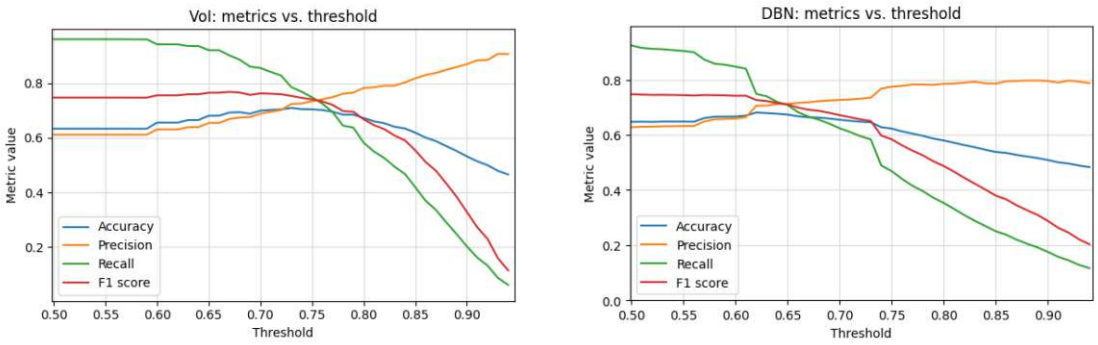
Figure 6.4: Trajectory comparison for patient `hadm_id`=29679030

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

6.0.5 Threshold Variation and Confidence Analysis

Since the performance metrics were based on predictions with confidence scores of at least 0.5, an analysis was conducted to assess whether the VoI approach improved prediction confidence. Lead time and accuracy were measured across a range of thresholds for both the VoI-enhanced DBN and the SOFA score.

Figure 6.5 shows the change in accuracy, precision, recall, and F1-Score with varying thresholds for both the VoI-enhanced DBN and the standard DBN. Compared to the performance of the DBN, the VoI approach maintains consistently higher metrics at elevated thresholds. When compared to SOFA scores, the VoI's performance begins to decline only above a threshold of 0.75, whereas the standard DBN metrics already deteriorate from 0.63.



(a) VoI performance metrics vs. threshold (b) Sepsis specific metrics vs. threshold

Figure 6.5: Comparison of VoI and DBN performance metrics across thresholds.

As the diagnostic threshold increases, the VoI-enhanced DBN maintains higher confidence than the standard DBN. The baseline DBN predicted sepsis earlier than SOFA only up to a threshold of 0.72, whereas the VoI approach extends this advantage to 0.85 (see Figure 6.6). This improved robustness under stricter criteria yields more significant early-diagnosis benefits. Even at slightly lower thresholds, VoI predicts sepsis earlier in approximately 70% of cases.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

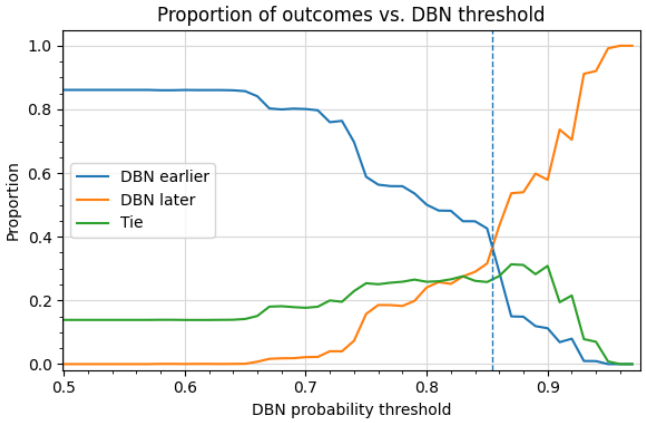


Figure 6.6: Proportion of outcomes vs. DBN threshold (VoI approach)

Figure 6.7 presents a heatmap of early-prediction lead times across varying thresholds for both the VoI-enhanced DBN and SOFA scores. It can be seen that even at a low SOFA severity of 2, the VoI approach leads in early detection across most confidence thresholds. Under stricter diagnostic criteria, VoI remains significantly earlier, and for high-risk cases (SOFA scores 6–8), the VoI method outperforms at nearly every threshold. This demonstrates a robust advantage of VoI in achieving earlier sepsis diagnosis over a wide range of operating points.

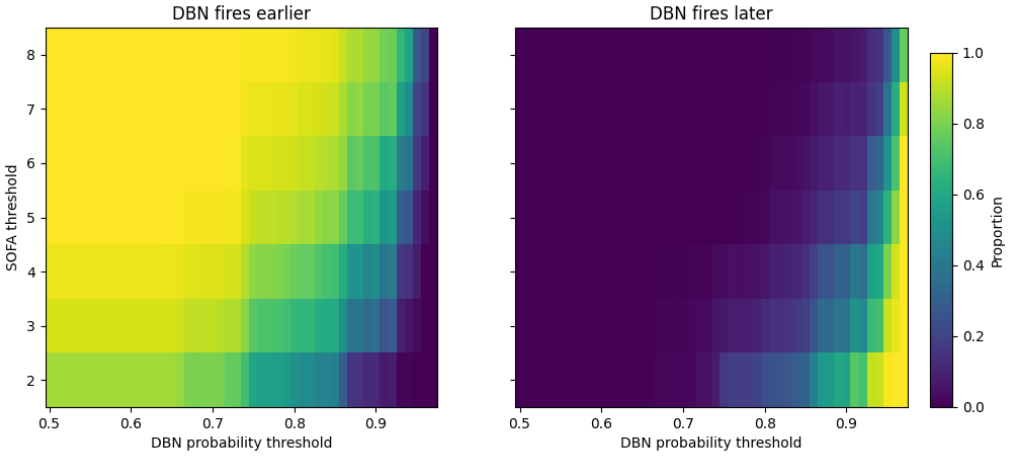


Figure 6.7: Heat mMap of DBN vs. SOFA thresholds (VoI approach)

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

CHAPTER 7

Conclusion

This thesis explored whether a constrained Dynamic Bayesian Network can anticipate sepsis earlier than the SOFA rule while preserving interpretability and accurate predictions. Structure and conditional probability tables of the model are estimated from historical data under clinical constraints (forward-only causality with sepsis as the latent parent within each slice). After training, these fixed parameters are applied to incoming laboratory results to compute posterior sepsis probabilities over time. On a balanced evaluation data set the DBN matched SOFA on overall accuracy and F1, reduced missed cases at comparable operating points, and issued sepsis alerts earlier. This indicates a clinically meaningful advantage in the ICU, where hours can alter outcomes.

To complement the DBN's sequential inference at prediction time, a Value-of-Information framework was introduced to select the next laboratory test that is expected to increase diagnostic confidence the most. Candidate measurements are ranked by their expected reduction in posterior uncertainty given the current evidence, by computing the entropy. This results in an ordering of tests that widen the early warning window while maintaining calibration across decision thresholds. The contribution is therefore not only a discriminative temporal model, but also an indicator selecting which test to draw next, allowing clinicians to balance timeliness against certainty.

On the 30% hold out test set (about 2,700 non-septic and 3,508 septic ICU stays), the DBN with a probability threshold of 0.5 achieved an accuracy of 0.645, precision of 0.628, recall of 0.917, and an F1-score of 0.746. SOFA (≥ 2) reached an accuracy of 0.641, precision of 0.630, recall of 0.885, and an F1-score of 0.736. The confusion matrices show that the DBN produced fewer false negatives than SOFA (290 vs. 403), which means fewer missed septic patients. In terms of timing, the DBN predicted sepsis earlier in 2,078 cases (SOFA earlier in 157; tied in 649). The median lead was about two lab results (mean 1.6 labs). Converted to time, this corresponds to a median lead of roughly 0.8 hours and a mean lead of about 4.2 hours. These gains are small in appearance

but material in practice, because even short advances can bring forward antibiotics and monitoring.

Adding the Value-of-Information approach improved early detection further. With VoI, sepsis was predicted earlier in 2,606 cases, the mean lead increased to about two labs while the median stayed at two labs, and the time lead rose to a median of 1.5 hours and a mean of 7.1 hours. VoI also made the model more robust under varying probability thresholds. The DBN was earlier than SOFA up to a threshold of about 0.72 without VoI, and up to about 0.85 with VoI. By selecting the most informative first lab, the VoI approach raised diagnostic confidence quickly, often reaching the alert threshold after fewer measurements than SOFA. Together, these results show that a fixed, offline-trained DBN can deliver earlier warnings than SOFA, and that VoI can widen this early-warning window while keeping probability estimates reliable across thresholds.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

List of Figures

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

List of Tables

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

List of Algorithms

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Full ICD-9 to ICD-10 Sepsis Code Mappings

Full ICD-9 to ICD-10 Sepsis Code Mappings

Table 1: Complete CCS Septicemia to ICD-9 and ICD-10 sepsis code mappings

ICD-9	ICD-10
0031 Salmonella septicemia	A021 Salmonella sepsis
0202 Septicemic plague	A207 Septicemic plague
0223 Anthrax septicemia	A227 Anthrax sepsis
0362 Meningococcemia	A392 Acute meningococcemia; A393 Chronic meningococcemia; A394 Meningococcemia, unspecified
0380 Streptococcal septicemia	A400 Sepsis due to Streptococcus, group A; A401 Sepsis due to Streptococcus, group B
0381 Staphylococc septicemia (pre-1997)	A411 Sepsis due to other specified Staphylococcus
03810 Staph septicemia, unspecified ('97+)	A412 Sepsis due to unspecified Staphylococcus
03811 Staph aureus septicemia ('97+)	A4101 Sepsis due to MSSA
03812 MRSA septicemia ('08+)	A4102 Sepsis due to MRSA
03819 Other staphylococcal septicemia	A411 Sepsis due to other specified Staphylococcus
0382 Pneumococcal septicemia	A403 Sepsis due to Streptococcus pneumoniae; A408 Other streptococcal sepsis
0383 Anaerobic septicemia	A414 Sepsis due to anaerobes
03840 Gram-negative septicemia NOS	A4150 Gram-negative sepsis, unspecified
03841 H. influenzae septicemia	A413 Sepsis due to Haemophilus influenzae
03842 E. coli septicemia	A4151 Sepsis due to Escherichia coli
03843 Pseudomonas septicemia	A4152 Sepsis due to Pseudomonas

Continued on next page

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Continued from previous page

ICD-9	ICD-10
03844 Serratia septicemia	A4153 Sepsis due to Serratia
03849 Gram-negative septicemia NEC	A4154 Sepsis due to Acinetobacter baumannii; A4159 Other Gram-negative sepsis
0388 Septicemia NEC	A4181 Sepsis due to Enterococcus; A4189 Other specified sepsis
0389 Septicemia NOS	A409 Streptococcal sepsis, unspecified; A419 Sepsis, unspecified organism
0545 Herpetic septicemia	B007 Disseminated herpesviral disease
449 Septic arterial embolism ('07+)	
77181 Neonatal septicemia [sepsis] ('02+)	P360–P369 Sepsis of newborn (group B; unspecified and other streptococci; Staphylococci; E. coli; anaerobes; other)
7907 Bacteremia NOS	
99591 SIRS due to infection w/o organ dysfunction	R6520 Severe sepsis without septic shock
99592 SIRS due to infection with organ dysfunction	R6521 Severe sepsis with septic shock
	A267 Erysipelothrix sepsis
	A327 Listerial sepsis
	A427 Actinomycotic sepsis
	A5486 Gonococcal sepsis
	B377 Candidal sepsis
	O0337 Sepsis following incomplete spontaneous abortion
	O0387 Sepsis following complete/unspecified spontaneous abortion
	O0487 Sepsis following (induced) termination of pregnancy
	O0737 Sepsis following failed attempted termination of pregnancy
	O0882 Sepsis following ectopic and molar pregnancy
	O85 Puerperal sepsis
	O8604 Sepsis following an obstetrical procedure
	T8112XA Postprocedural septic shock, initial encounter
	T8144XA Sepsis following a procedure, initial encounter

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Bibliography

[1] PhysioNet. MIMIC-IV version 3.1. PhysioNet, 2024. URL https://physionet.org/content/mimiciv/3.1/.

[2] A. E. W. Johnson, L. Bulgarelli, L. Shen, A. Gayles, A. Shammout, S. Horng, T. J. Pollard, S. Hao, B. Moody, B. Gow, L. H. Lehman, L. A. Celi, and R. G. Mark. MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data*, 10(1):1, 2023. doi: 10.1038/s41597-022-01899-x.

[3] Lars Hempel, Sina Sadeghi, and Toralf Kirsten. Prediction of Intensive Care Unit length of stay in the MIMIC-IV dataset. *Applied Sciences*, 13(12):6930, 2023. doi: 10.3390/app13126930. URL https://www.mdpi.com/2076-3417/13/12/6930.

[4] Milad Naeimaei Aali, Felix Mannhardt, and Pieter Jelle Toussaint. MIMIC-IV-Ext-CEKG: A process-oriented dataset derived from MIMIC-IV for enhanced clinical insights (version 1.0.0), April 2025. URL https://physionet.org/content/mimic-iv-ext-cekg/1.0.0/. PhysioNet dataset.

[5] MIT Laboratory for Computational Physiology. MIMIC-IV v3.1 clinical database, 2024. URL https://physionet.org/content/mimiciv/3.1/. Access policy: Only credentialed users who sign the PhysioNet Data Use Agreement; Required training: CITI Data or Specimens Only Research.

[6] MIT Laboratory for Computational Physiology. Diagnosis timestamp (issue #918). GitHub Issue, MIT-LCP/mimic-code, May 2021. URL https://github.com/MIT-LCP/mimic-code/issues/918. Clarifies that diagnoses_icd has no per-diagnosis timestamp.

[7] Healthcare Cost and Utilization Project (HCUP). *CCS Category Names (Full Labels): Single-Level CCS Diagnosis Category Labels.* Agency for Healthcare Research and Quality, 2015. Accessed January 2024.

[8] Healthcare Cost and Utilization Project (HCUP). *Clinical Classifications Software Refined (CCSR) for ICD-10-CM Diagnoses User Guide, v2024.1.* Agency for Healthcare Research and Quality, Dec 2023.

[9] Chanu Rhee, Sameer S. Kadri, Susan S. Huang, Michael V. Murphy, Lingling Li, Richard Platt, and Michael Klompas. Objective sepsis surveillance using electronic clinical data. *Infection Control & Hospital Epidemiology*, 37(2):163–171, 2016. doi: 10.1017/ice.2015.264. Epub 2015 Nov 3.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

[10] Jean-Louis Vincent, Rafael Moreno, Jukka Takala, Stuart Willatts, Anabela De Mendonça, Herman Bruining, C. Kathryn Reinhart, Peter M. Suter, and Luc G. Thijs. The SOFA (Sepsis-Related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Medicine*, 22(7):707–710, 1996. doi: 10.1007/BF01709751.

[11] Nir Friedman, Kevin P. Murphy, and Stuart J. Russell. Learning the structure of dynamic probabilistic networks. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 139–147, 1998.

[12] Thomas Dean and Keiji Kanazawa. A model for reasoning about persistence and causation. *Computational Intelligence*, 5(3):142–150, 1989. doi: 10.1111/j.1467-8640.1989.tb00324.x.

[13] Kevin Patrick Murphy. *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD thesis, University of California, Berkeley, 2002.

[14] José A. Gámez, Juan L. Mateo, and José M. Puerta. Learning Bayesian networks by hill climbing: Efficient methods based on progressive restriction of the neighborhood. *Data Mining and Knowledge Discovery*, 22(1–2):106–148, 2011.

[15] Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2): 461–464, 1978.

[16] Remco Ronaldus Bouckaert. *Bayesian Belief Networks: From Construction to Inference*. PhD thesis, Utrecht University, 1995.

[17] Nicolai Meinshausen and Peter Bühlmann. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):417–473, 2010.

[18] Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, Cambridge, MA, 2009.

[19] David Heckerman, Dan Geiger, and David M. Chickering. Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20(3):197–243, 1995.

[20] Ankur Ankan and Abinash Panda. pgmpy: Probabilistic graphical models using Python. In *Proceedings of the 14th Python in Science Conference (SciPy)*, pages 6–11, 2015.

[21] Stanley F. Chen and Joshua Goodman. An empirical study of smoothing techniques for language modeling. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 310–318, 1996.

[22] Elisabeth Fenwick, Lotte Steuten, Saskia Knies, Salah Ghabri, Anirban Basu, James F. Murray, Hendrik E. Koffijberg, Mark Strong, Gillian D. Sanders Schmidler, and Claire Rothery. Value of information analysis for research decisions—an introduction. *Value in Health*, 23(2):139–150, 2020.

[23] Andreas Krause and Carlos Guestrin. Near-optimal nonmyopic value of information in graphical models. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 324–331, 2005.

[24] Senthil K. Nachimuthu and Peter J. Haug. Early detection of sepsis in the emergency department using dynamic Bayesian networks. In *Proceedings of the AMIA Annual Symposium*, pages 653–662, 2012.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

[25] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[26] Tony Wang, Tom Velez, Emilia Apostolova, Tim Tschampel, Thuy L. Ngo, and Joy Hardison. Semantically enhanced dynamic Bayesian network for detecting sepsis mortality risk in ICU patients with infection. *arXiv preprint arXiv:1806.10174*, 2018.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.