

공학박사 학위논문

**A Continuous Vital Sign Analysis
Methodology for Identifying Novel
Physiological Markers in Neonatal
Intensive Care Unit**

신생아 중환자실 임상 예후 향상을 위한 연속
활력징후 기반 생리 지표 분석 방법론 연구

2025년 7월

서울대학교 융합과학기술대학원

헬스케어융합학과

송 원 근

A Continuous Vital Sign Analysis Methodology for Identifying Novel Physiological Markers in Neonatal Intensive Care Unit

지도 교수 이 교 구

이 논문을 공학박사 학위논문으로 제출함
2025년 7월

서울대학교 융합과학기술대학원
헬스케어융합학과
송 원 근

송원근의 공학박사 학위논문을 인준함
2025년 7월

위 원 장	<u>최 창 원</u>	(인)
부위원장	<u>이 교 구</u>	(인)
위 원	<u>유 수 영</u>	(인)
위 원	<u>이 호 영</u>	(인)
위 원	<u>고 태 훈</u>	(인)

Abstract

Low birth weight preterm infants admitted to neonatal intensive care units (NICUs) represent a high-risk group with high mortality and morbidity rates. These preterm infants require continuous physiological monitoring and intensive clinical intervention. Early diagnosis and prognosis prediction are critical for improving survival and long-term outcomes in preterm infants. To address this need and support timely clinical decisions, recent research has extensively focused on developing predictive models and identifying clinical indicators from continuous vital sign data.

However, machine learning and deep learning models applied to NICU preterm infant data frequently fail to demonstrate statistically significant superiority over logistic regression models, often exhibiting suboptimal performance during external validation. These limitations result from several challenges. First, institutional and research-specific variations in physiological signal acquisition and processing methodologies impede the generalizability of predictive models. Second, considerable heterogeneity in gestational age and the frequency of clinical interventions across institutions and care providers complicates the extraction of stable, reliable indicators. Furthermore, studies using continuous vital signs data in preterm infants are limited by considerable constraints due to high computational burdens and the restricted applicability of time series analysis methods.

This study proposed a scalable methodology for continuous vital sign analysis to address analytical complexities resulting from preterm infant characteristics, computational resource demands, constraints in time series analysis application, and existing research limitations. Our methodology efficiently integrated diverse time series analysis methods from various research domains. This enables the identification of clinically relevant diagnostic and prognostic factors from continuously acquired large-scale vital sign data and supports in-depth exploration

of novel physiological factors. We developed a scalable feature extraction approach to derive previously uncharacterized continuous vital sign-based features, applying time series analysis methods to capture dynamic features reflective of NICU-specific physiological patterns often challenging to extract from electronic medical record (EMR) data. Additionally, by transforming advanced false discovery rate (FDR) control and clinical trial emulation methods into partitionable algorithms, the proposed methods improved both the scalability and robustness of identified clinical indicators. These methods, by enabling parallel and distributed computing, substantially enhance computational efficiency and overall scalability for large-scale multicenter clinical studies, aligning with current high-performance computing paradigms.

To validate the proposed methodology, we conducted several studies using the methods that were implemented by the framework. We initially performed simulation analyses to determine that our proposed FDR control method provides superior control and computational efficiency compared to traditional methodologies. Subsequently, we developed predictive models for sepsis and mortality, critical complications in preterm infants, based on the proposed framework. These models demonstrated robust classification performance even in external validation datasets. We further validated that continuous physiological signal-based predictive models, developed using the proposed framework, can contribute to clinical decision-making. Lastly, by identifying the novel predictor for intraventricular hemorrhage (IVH) via time series analysis methods from other research domains, we demonstrated the capacity of the proposed methodology to discover new clinical indicators.

This study provides several notable contributions. We systematically derived high-resolution clinical indicators from continuous vital sign data, thereby expanding the scope and precision of feature extraction and selection methodologies

for risk factor identification and prediction in preterm infants. We also developed a time series analysis framework that accurately reflects the physiological characteristics of preterm infants, consequently mitigating limitations inherent in existing continuous vital sign analysis methodologies for this population. Furthermore, through external validation of models developed using our proposed methodology, we enhanced the reliability and reproducibility of predictive models within the NICU. Finally, via in-depth analysis of novel physiological predictors, we aimed to enhance model interpretability and clinical utility by establishing links between physiological characteristics and significant clinical symptoms, such as autonomic nervous system dysfunction. Overall, this study addresses existing computational and analytical constraints, thereby improving the practical applicability of continuous vital signs analysis research. We anticipate our methodology will support enhanced research convenience and facilitate the resolution of critical clinical questions. Moreover, it is expected to advance prognostic assessment in preterm infants and contribute to the development of dependable and clinically actionable artificial intelligence models within the NICU environment.

Keyword: Preterm infants, Continuous vital signs, FDR control, Time series analysis, High-dimensional data, Parallel computing, Target trial emulation

Student Number: 2022-37947

Table of Contents

Abstract	i
Table of Contents	iv
List of Tables	vii
List of Figures	ix
Chapter 1. Introduction	1
1.1. Clinical Background: Low Birth Weight Preterm Infants in a Neonatal Intensive Care Unit	1
1.1.1. Preterm and Low Birth Weight Infants.....	1
1.1.2. Mortality and Complications of Preterm Infants in NICU	2
1.1.3. Early Detection in the NICU: Importance and Challenges	4
1.2. Analytic Methods and Frameworks for Preterm Infants in NICU.....	5
1.2.1. ECG Based Approach.....	6
1.2.2. EMR-based Approaches	7
1.2.3. Continuous Vital signs-based Approaches	9
1.2.4. Feature Selection Methods for High-Dimensional Data.....	10
1.3. Research Questions and Aims.....	13
Chapter 2. Continuous Vital Signs Derived Feature Extraction and Analytic Frameworks for Identifying Risk Factors in Preterm Infants	16
2.1. Introduction.....	16
2.2. Methods	18
2.2.1. Design Principles	18
2.2.2. Time Series Analysis-based Feature Extraction Approach.....	20
2.2.3. Continuous Vital Signs Feature Calculation Methods	20
2.2.4. Continuous Vital Signs Feature Analysis and Selection.....	23
2.2.5. Data Analysis Framework Implementation	31

2.2.6. Validation Strategies.....	37
2.3. Results.....	38
2.3.1. Evaluation of Parallel Procedure in Execution Time.....	38
2.3.2. Evaluation of FDR and Sensitivity for relevant features	41
Chapter 3. Development and Validation of a Predictive Model for Clinical Critical Events in Preterm Infants Admitted to the NICU.....	44
3.1. Introduction.....	44
3.2. Methods	44
3.2.1. Study Design.....	44
3.2.2. Data sources.....	45
3.2.3. Eligibility criteria and outcome	46
3.2.4. Predictors	46
3.2.5. Statistical Analysis.....	47
3.2.6. Predictive Model Development and Evaluation	47
3.3. Results.....	48
3.3.1. Study Population.....	48
3.3.2. Identified Features of Clinical Critical Events.....	51
3.3.1. Predictive Models Performance.....	57
3.4. Discussion.....	59
Chapter 4. Predictive Modeling of Extubation Readiness in Preterm Infants Using Real-Time Physiological Data.....	62
4.1. Introduction.....	62
4.2. Methods	63
4.2.1. Study Design.....	63
4.2.2. Eligibility Criteria and Outcome.....	63
4.2.3. Predictors	64
4.2.4. Statistical Analysis.....	66
4.2.5. Predictive Model Development and Evaluation	67
4.3. Results.....	69
4.3.1. Study Population.....	69

4.3.2. Predictors of Extubation Failure	71
4.3.3. Model Performance.....	72
4.4. Discussion.....	73
Chapter 5. New physiological Risk Factor of Intraventricular Hemorrhage of Preterm Infant	76
5.1. Introduction.....	76
5.2. Methods	77
5.2.1. Study Design.....	77
5.2.2. Data sources.....	77
5.2.3. Case and Control Definition	78
5.2.4. Covariates	78
5.2.5. Statistical Analysis.....	79
5.3. Results.....	80
5.3.1. Study Design.....	80
5.3.2. Risk Factors Associated with IVH.....	83
5.3.3. Oxygen Saturation Decorrelation Time	83
5.3.4. Sensitivity Analyses.....	89
5.4. Discussion.....	90
Chapter 6. Conclusion	95
Bibliography.....	99
Abstract in Korean	111

List of Tables

TABLE 3-1. BASELINE CHARACTERISTICS	51
TABLE 3-2. SELECTED FEATURES FOR ALL-CAUSE MORTALITY FROM THE DEVELOPMENT COHORT.	52
TABLE 3-3. SELECTED FEATURES FOR ALL-CAUSE MORTALITY FROM THE EXTERNAL VALIDATION COHORT.	53
TABLE 3-4. SELECTED FEATURES FOR LATE ONSET SEPSIS FROM THE DEVELOPMENT COHORT.	54
TABLE 3-5 PERFORMANCE OF PREDICTIVE MODELS USING THE PROPOSED CONTINUOUS VITAL SIGN ANALYSIS METHOD FRAMEWORKS FOR CLINICAL CRITICAL EVENTS.	57
TABLE 3-6 PERFORMANCE OF PREDICTIVE MODELS USING THE PROPOSED CONTINUOUS VITAL SIGN ANALYSIS METHOD FRAMEWORKS FOR ALL-CAUSE MORTALITY.	58
TABLE 3-7 PERFORMANCE OF REAL-TIME PREDICTIVE MODELS USING THE PROPOSED CONTINUOUS VITAL SIGN ANALYSIS METHOD FRAMEWORKS FOR CLINICAL CRITICAL EVENTS.....	58
TABLE 3-8 PERFORMANCE OF REAL-TIME PREDICTIVE MODELS USING THE PROPOSED CONTINUOUS VITAL SIGN ANALYSIS METHOD FRAMEWORKS FOR ALL-CAUSE MORTALITY.	58
TABLE 4-1. FEATURE EXTRACTION METHOD USED IN NEXt PREDICTOR MODEL [131].....	65
TABLE 4-2 HYPERPARAMETERS FOR DEVELOPING PREDICTION MODELS	68
TABLE 4-3 BASELINE CHARACTERISTICS IN THE DEVELOPMENT COHORT	70
TABLE 4-4 BASELINE CHARACTERISTICS IN THE INTERNAL VALIDATION COHORT.....	71
TABLE 4-5. CLINICALLY RELEVANT FEATURES IDENTIFIED FOR EXTUBATION.	72
TABLE 4-6 PERFORMANCE OF PREDICTIVE MODELS UTILIZING THE PROPOSED CONTINUOUS VITAL SIGN ANALYSIS METHODOLOGY.	72
TABLE 5-1 BASELINE CHARACTERISTICS.....	82
TABLE 5-2 UNIVARIATE AND MULTIVARIABLE ANALYSIS OF RISK FACTORS ASSOCIATED WITH IVH.....	83

TABLE 5-3. RATIOS OF RISK FACTORS FOR EACH GROUPS.	90
TABLE 5-4. ODDS RATIOS OF RISK FACTORS FOR VARIOUS REGRESSION MODELS.....	90

List of Figures

FIGURE 2-1. PROPOSED FEATURE ANALYSIS AND SELECTION METHOD.....	24
FIGURE 2-2 PROPOSED CONTINUOUS VITAL SIGNS ANALYSIS PROCESS.	32
FIGURE 2-3. MAPREDUCE MODEL [109].....	33
FIGURE 2-4 MAP-REDUCE WORKFLOW FOR SUBSAMPLING AND FEATURE EXTRACTION.....	34
FIGURE 2-5 INTEGRATED WORKFLOW FOR EMULATING CASE-CONTROL STUDIES AND FEATURE SELECTION.....	36
FIGURE 2-6 COMPARISON OF EXECUTION TIMES FOR PARALLEL AND SEQUENTIAL PROCEDURES ACROSS VARYING FEATURE DIMENSIONS.	40
FIGURE 2-7 EXECUTION TIME BY SAMPLE SIZE AND THE NUMBER OF FEATURES: (A) PARALLEL PROCEDURE, (B) SEQUENTIAL PROCEDURE.....	40
FIGURE 2-8 EXECUTION TIME BY SAMPLE SIZE AND THE NUMBER OF FEATURES IN PARALLEL PROCEDURE.	41
FIGURE 2-9 SENSITIVITY RESULTS FROM THE SIMULATION STUDY.....	43
FIGURE 2-10 FALSE DISCOVERY RATE RESULTS FROM THE SIMULATION STUDY.	43
FIGURE 3-1. DEVELOPMENT AND INTERNAL AND EXTERNAL VALIDATION COHORT TO DEVELOP AND EVALUATE ALL-CAUSE PREDICTIVE MODELS. (A) DEVELOPMENT COHORT AND INTERNAL VALIDATION COHORT. (B) EXTERNAL VALIDATION COHORT.	49
FIGURE 3-2 ABSOLUTE CORRELATION MATRIX DIAGRAM IN ALL-CAUSE MORTALITY (A) FEATURES CLUSTERING DENDROGRAM, WITH LEAVES REPRESENTING INDIVIDUAL FEATURES AND NODES INDICATING CLUSTERS. (B) FEATURES GROUPED BY VITAL SIGN TYPE.....	55
FIGURE 3-3 ABSOLUTE CORRELATION MATRIX DIAGRAM OF SELECTED FEATURE IN ALL- CAUSE MORTALITY.	55
FIGURE 3-4 ABSOLUTE CORRELATION MATRIX IN LATE ONSET SEPSIS. (A) FEATURES CLUSTERING DENDROGRAM, WITH LEAVES REPRESENTING INDIVIDUAL FEATURES AND NODES INDICATING CLUSTERS. (B) FEATURES GROUPED BY VITAL SIGN TYPE.....	56
FIGURE 3-5 ABSOLUTE CORRELATION DIAGRAM OF SELECTED FEATURES IN LATE ONSET	

SEPSIS.....	56
FIGURE 3-6 CLINICAL CRITICAL EVENT PREDICTIVE MODELS AUROC PERFORMANCE: (A) INTERNAL VALIDATION COHORT, (B) EXTERNAL VALIDATION COHORT.....	59
FIGURE 3-7 REAL-TIME CLINICAL CRITICAL EVENT PREDICTIVE MODELS AUROC PERFORMANCE (A) INTERNAL VALIDATION COHORT, (B) EXTERNAL VALIDATION COHORT.....	59
FIGURE 4-1. DEVELOPMENT AND INTERNAL VALIDATION COHORT TO IDENTIFY PREDICTORS AND DEVELOP PREDICTIVE MODELS.....	69
FIGURE 5-1 FLOW DIAGRAM OF INCLUSIONS AND EXCLUSIONS FOR THE STUDY.....	80
FIGURE 5-2 TREND AND INSTABILITY DENSITY IN SpO ₂ DECORRELATION TIME. (A) LINE PLOT OF SpO ₂ DECORRELATION TIME OVER THE FIRST 24 H AFTER BIRTH IN THE IVH AND NON-IVH GROUPS. THE SHADED AREA SHOWS 95% CI OF THE SpO ₂ DECORRELATION TIME. (B) LINE PLOT OF RAW SpO ₂ OVER THE FIRST 24 H AFTER BIRTH IN THE IVH AND NON-IVH GROUPS. THE SHADED AREA REPRESENTS THE 95% CI OF SpO ₂ . (C) DENSITY PLOT OF AUTOCORRELATION BY TIME LAG IN THE IVH AND NON-IVH GROUPS. THE X-AXIS REPRESENTS THE ELAPSED HOURS AFTER BIRTH, THE Y-AXIS REPRESENTS THE TIME LAG, AND THE COLOR INTENSITY INDICATES THE PROJECTIONS OF AUTOCORRELATION VALUES BY TIME LAG.....	85
FIGURE 5-3. TREND AND DENSITY PLOT OF SpO ₂ DECORRELATION TIME ACROSS GROUPS IN THE FIRST 24 HOURS AFTER BIRTH.....	86
FIGURE 5-4. TREND AND DENSITY PLOT OF SpO ₂ DECORRELATION TIME IN EACH ORDER DIFFERENTIAL. (A) LINE PLOT OF SpO ₂ DECORRELATION TIME OVER THE FIRST 24 H AFTER BIRTH IN THE IVH AND NON-IVH GROUPS. (B) DENSITY PLOT OF AUTOCORRELATION BY TIME LAG IN THE IVH AND NON-IVH GROUPS. (C) LINE PLOT OF FIRST-ORDER DIFFERENTIAL SpO ₂ DECORRELATION TIME OVER THE FIRST 24 H AFTER BIRTH IN THE IVH AND NON-IVH GROUPS. (D) DENSITY PLOT OF FIRST-ORDER DIFFERENTIAL SpO ₂ AUTOCORRELATION BY TIME LAG IN THE IVH AND NON-IVH GROUPS. (E) LINE PLOT OF SECOND-ORDER DIFFERENTIAL SpO ₂ DECORRELATION TIME OVER THE FIRST 24 H AFTER BIRTH IN THE IVH AND NON-IVH GROUPS. (F) DENSITY	

PLOT OF SECOND-ORDER DIFFERENTIAL SpO_2 AUTOCORRELATION BY TIME LAG IN THE IVH AND NON-IVH GROUPS.....	87
FIGURE 5-5. AUTOCORRELATION DENSITY PLOT OF SpO_2 DECORRELATION TIME AT 18 HOURS POST-BIRTH. (A) DENSITY PLOT OF AUTOCORRELATION IN THE IVH AND NON-IVH GROUPS. (B) DENSITY PLOT OF AUTOCORRELATION BY TIME LAG IN THE IVH WITH GRADE I, II, III AND IV GROUPS.	
	88
FIGURE 5-6. FOREST PLOT OF SpO_2 DECORRELATION TIME FOR EACH SAMPLING PERIOD.	89

Chapter 1. Introduction

1.1. Clinical Background: Low Birth Weight Preterm Infants in a Neonatal Intensive Care Unit

1.1.1. Preterm and Low Birth Weight Infants

Preterm birth, defined as delivery before 37 weeks of gestation age (GA), remains a major global health challenge and a leading cause of neonatal morbidity and mortality [1]. Infants born preterm are physiologically immature and clinically vulnerable, with a heightened risk of developing life-threatening complications due to underdeveloped organ systems.

Preterm infants require substantial medical support to survive in the extrauterine environment due to the immaturity of multiple organ systems, including the lungs, brain, cardiovascular system, and gastrointestinal tract. In utero, fetal organs undergo progressive maturation to achieve functional competence necessary for extrauterine life. In term neonates, this maturation typically enables spontaneous respiration, effective pulmonary gas exchange, metabolic homeostasis, autonomic regulation of cardiovascular function, neurologic responsiveness to sensory stimuli, coordinated gastrointestinal motility with enzymatic activity, and the presence of primitive reflexes such as sucking, grasping, and rooting [1, 2].

Preterm infants with a higher degree of immaturity are associated with physiological vulnerability, leading to a steep increase in the risk of both mortality and short- and long-term morbidity. Therefore, delivery decisions and treatment

strategies, including initiation of active neonatal treatment, or timing and intensity of interventions, must be guided by the infant's level of maturity, underscoring the need for validated clinical indices and objective measures to assess maturity.

GA has been the most used indicator and serves as a practical proxy measure for developmental maturity. Preterm infants were defined as belonging to one of three gestational age categories: extremely preterm (less than 28 weeks), very preterm (28 to 31 weeks), and moderately preterm (32 to 36 weeks) [1, 3]. These classifications are associated with clinical outcomes, as survival rates and the incidence of complications [3]. However, GA based classifications alone do not fully capture the spectrum of neonatal immaturity. Specifically, GA does not determine for infants who are born too early versus those who are small for gestational age (SGA), nor does it address cases of functional immaturity in full-term infants. For this reason, additional proxy measures are required to complement GA when evaluating neonatal maturity.

1.1.2. Mortality and Complications of Preterm Infants in NICU

Recent advances in medical technology and ongoing research have led to significant progress in perinatal and neonatal intensive care, resulting in substantial improvements in overall survival rates [4-6]. Nevertheless, even at present, low birth weight preterm infants have higher risks of morbidity and mortality compared to term infants. Globally, the incidence of preterm birth has remained relatively stable (9.8% in 2010 and 9.9% in 2020). As of 2019, mortality related to preterm birth complications accounted for 17.7% of all neonatal deaths worldwide [7, 8].

Compared with outborn pre-term infants, those admitted to NICUs generally have significantly higher survival rates. Consequently, the spectrum and causes of

death differ substantially between the overall population of preterm infants and those who receive intensive care. In addition, there is considerable variation in neonatal outcomes depending on country, institutional NICU level, available medical resources, and the quality of perinatal and neonatal care. These disparities are further compounded by differences in clinical definitions, diagnostic thresholds, and reporting standards for neonatal complications and causes of death.

The primary complications and causes of death for preterm infants in the NICU vary slightly by region and institution. However, respiratory failure, infection, and neurological injury are consistently identified complications.

In the United States, a study of extremely preterm infants admitted to NICUs between 2013 and 2018 reported an approximate 20% incidence of late-onset sepsis (LONS) in infants born at 22–28 weeks' gestational age. During the same period, bronchopulmonary dysplasia (BPD) occurred in 8.0%, intracranial hemorrhage in 14.3%, and neurodevelopmental impairments affected 29.3% (moderate) and 21.2% (severe) of these infants. While survival rates significantly increased compared to the 2008–2012 period, the incidence of neurodevelopmental impairments remained unchanged [9].

Due to the heterogeneity in NICUs, clinical practice, and data classification systems, it remains difficult to establish universally accepted definitions for major complications of prematurity. For instance, diagnostic criteria for conditions such as intraventricular hemorrhage (IVH), BPD, necrotizing enterocolitis (NEC), and sepsis may vary across institutions and countries in terms of imaging modality, timing of diagnosis, and clinical thresholds for intervention. Moreover, differences in gestational age thresholds for viability and in the ethical framework guiding resuscitation and intensive care practices influence the clinical course and reported outcomes. As a result, international comparisons of morbidity and mortality data are

often limited by inconsistency in terminology and methodology.

This variability significantly impacts the ability to generate standardized epidemiologic profiles of preterm complications, hindering efforts to implement globally harmonized quality improvement initiatives and evidence-based policy recommendations. Therefore, the interpretation of both short- and long-term outcomes in NICU-admitted preterm infants must be contextualized within the specific healthcare environment, gestational age distribution, and national clinical guidelines.

1.1.3. Early Detection in the NICU: Importance and Challenges

Early detection and timely intervention during clinical deterioration can significantly improve the prognosis and survival rates of preterm infants. There is growing evidence and research supporting that early diagnosis and detection of patient deterioration in the NICU can prevent severe complications and improve clinical outcomes, ultimately leading to further improvements in preterm infant survival rates.

One of the primary subjects of research focused on early detection is late-onset sepsis. Sepsis is a complication that arises from infection and remains one of the most severe complications. Despite advances in neonatal care, sepsis remains a leading contributor to morbidity and mortality in NICUs. It has been reported that approximately 20% of deaths among infants weighing less than 1,500g were attributable to sepsis. Moreover, the risk of death is nearly threefold higher in infants diagnosed with sepsis compared to those without infection [10]. Neonatal sepsis is typically classified into two categories: early-onset sepsis (EOS), occurring within the first 72 hours of life, and LONS, which presents between 72 hours and 120 days

after birth [10, 11]. EOS is primarily the result of intrauterine infection or vertical transmission of pathogens during labor and delivery, whereas LONS may arise from both vertical transmission and horizontal acquisition of bacteria from healthcare personnel or the NICU environment.

EOS can be prevented in up to 80% of cases through the administration of intrapartum antibiotic prophylaxis [12]. In contrast, no established prophylactic strategy or standardized guidelines currently exist for early identification of LONS, particularly in asymptomatic infants. The definitive diagnosis of LONS requires blood culture testing; however, results typically take 48 to 72 hours. To avoid treatment delays, empirical antibiotic therapy is often initiated before confirmation. Even when blood cultures are negative, antibiotics are frequently continued if clinical symptoms of LONS are present, due to the potential for false-negative culture results. However, this approach results in antimicrobial resistance, exposing infants to the risks associated with prolonged antibiotic use, and increases healthcare costs. Furthermore, preterm infants have a limited blood volume available for sampling, and blood culture is an invasive procedure with a high rate of false-negative results.

1.2. Analytic Methods and Frameworks for Preterm Infants in NICU

Predictive tools and analytical models have been developed to support early recognition of clinical deterioration and to guide prognostic assessment and timely intervention in preterm infants, beginning at the time of delivery and extending throughout the NICU stay. These models vary in clinical objective, timing of

assessment, and types of data utilized. A substantial number of vital signs-based approaches rely on continuous monitoring of vital signs, most derived from electrocardiograms (ECGs) and linked with electronic medical records (EMRs), to detect subtle physiological changes that result from physiological instability. However, the high computational demands and specialized equipment required for ECG-based approaches limit their applicability across all healthcare settings. Additionally, EMR data are often recorded at low temporal resolution and may be influenced by clinician-driven documentation bias. In response to these limitations, an increasing number of studies investigate the use of continuous vital sign data obtained from patient monitoring systems as a complementary approach.

1.2.1. ECG Based Approach

The ECG is a fundamental diagnostic tool that measures the heart's electrical activity during cardiac cycles. ECGs are recorded by processing and amplifying depolarization and repolarization signals, derived from the differential voltage between two points relative to a single ground reference. Standard 12-lead ECGs measure these differential voltages from various angles around the heart.

While the typical adult QRS complex duration ranges from 60 to 100 milliseconds, the neonatal QRS complex is comparatively shorter, spanning 30 to 94 milliseconds. Consequently, some characteristics and normal ranges of neonatal ECGs differ from those in adults. Furthermore, certain features may necessitate the analysis of high-frequency data in preterm infant population.

Heart rate variability (HRV) is one of the most widely used analytical methods. It quantifies the variation in RR intervals—or normal-to-normal intervals—between QRS complexes on ECG, which reflects autonomic nervous system function and maturation [13-15]. In particular, decreased HRV correlates with severe

inflammation and infection, and many studies have been therefore applied HRV as an analytical method for the early identification of LONS and NEC in neonates.[16, 17] Griffin et al[18, 19] presented the heart rate characteristics (HRC) index, which used heart rate variability and transient decelerations to identify early detection of neonatal sepsis. Furthermore, a multicenter, prospective, randomized controlled trial of the HeRO monitor, which is based on HRC analysis, demonstrated an approximate 22% reduction in all-cause mortality in the patient group that received HRC index [20]. Research consistently reports a strong association between vital sign instability in preterm infants and various adverse outcomes, including NEC, BPD, IVH, retinopathy of prematurity (ROP), cerebral palsy, and delayed early cognitive development. Specifically, HRV has shown a high correlation with these morbidities. [21-27]. However, these studies investigating the association between HRC index and morbidities other than LONS and NEC are often limited to single-center designs with small patient cohorts, requiring further research. Furthermore, existing ECG-based heart rate studies, including the HeRO system, demand substantial computational resources and often require specialized installation, hindering easy accessibility across multiple centers. An additional consideration for the HRC index is that its values can be influenced by external factors like surgical procedures and interventions, not solely a specific morbidity. Therefore, further research is needed to explore the relationship between HRC, other vital signs, and comprehensive clinical data.

1.2.2. EMR-based Approaches

The development of medical information systems and the widespread implementation of EMR have markedly increased the volume and accessibility of clinical and physiological data in NICUs. Research on preterm infants in the NICU

has often been constrained by small sample sizes and single-center designs, limiting external validation and broader applicability. The introduction of publicly accessible databases, such as MIMIC-III (Medical Information Mart for Intensive Care) [28], provides structured NICU EMR data that support validation efforts and enable multicenter analyses. As a result, research efforts utilizing these data sources have expanded, particularly in the application of artificial intelligence (AI) to detect and predict critical events and morbidities in preterm infants.

Consequently, many studies in this area have used EMR data due to its availability and the practical advantage of not requiring additional equipment. EMR data are typically available for retrospective analysis and are structured for algorithm development. However, several limitations in using EMR data for predictive modeling have been identified. One major limitation is the low temporal resolution of vital signs in EMR systems. In typical clinical workflows, vital signs may be recorded every 15 to 60 minutes, depending on institutional practices. This frequency is insufficient to capture the minute-to-minute physiological variability that can precede deterioration in preterm infants [29]. As a result, predictive models trained on EMR data may fail to detect early warning signs or subtle changes that are critical in a neonatal context [30].

Another issue is the subjectivity in EMR documentation. Data entries are often made at the discretion of the clinician, reflecting specific moments tied to care decisions or clinician awareness rather than continuous patient status [31, 32]. This introduces potential bias, making the data less representative of the full clinical picture and more influenced by provider behavior, workload, and institutional routines. Consequently, the validity of predictive algorithms based solely on EMR data may be limited.

Overall, while EMR-based approaches offer accessibility and convenience,

they are inherently limited by their temporal granularity and reliance on clinician input. Further research is needed to determine the most effective ways to leverage both data types in building robust, clinically useful predictive tools for neonatal care.

1.2.3. Continuous Vital signs-based Approaches

Although survival rates in NICUs have improved significantly over the past several decades, the information provided by conventional medical technologies have reached its limitations [33]. The limitations of traditional EMR data analysis have led to a significant increase in the perceived need for continuous vital sign analysis research since 2020 [27]. In response, there has been increasing interest in the use of continuous physiological data collected through bedside monitoring systems [27, 30, 34, 35]. These systems capture vital signs such as heart rate, respiratory rate, blood pressure, and oxygen saturation at high sampling rates, often on a second-by-second sampling period. This automated, human-independent data acquisition allows for higher temporal resolution and improved objectivity when compared to EMR records.

Recent studies suggested that incorporating continuous monitoring data into AI-based prediction models can improve the accuracy and timeliness of detecting adverse events, such as sepsis, apnea, and IVH [30, 36-38]. These findings imply a high potential for these approaches to support earlier intervention, particularly in high-risk neonatal populations.

Consequently, numerous studies are actively investigating features associated with severe morbidity and mortality in preterm infants. However, their implementation also raises new challenges. These include handling large volumes of time-series data, filtering signal noise, ensuring interoperability between different monitoring systems, and developing standards for integrating predictive outputs into

clinical decision-making processes. One of the most significant limitations is that most of these studies are unable to fully utilize observation windows exceeding one hour due to computational burden. Letzkus, et al. [39] investigated the association between low heart rate variability and cerebral palsy, developing a multivariable logistic regression classifier using 1 Hz heart rate from continuous vital signs. However, due to computational burden, they calculated features by extracting median values from 10-minute heart rate segments. Similarly, Niestroy, et al. [35] attempted to develop a model predicting all-cause mortality using numerous features generated from randomly extracted values within 10-minute segments for reducing computational burden. Peng, et al. [34] also generated features from continuous vital signs, similar to other studies, at a 0.5 Hz sampling rate. Yet, instead of utilizing observation windows exceeding one hour, their approach involved segmenting data into 10-minute intervals, calculating features, and then applying a grand mean. Consequently, this implies two key concerns: the potential for missing nuanced physiological signals embedded in continuous preterm infant vital signs, and the resulting inter-study variability in quantified features due to inconsistent aggregation approaches [40].

1.2.4. Feature Selection Methods for High-Dimensional Data

Feature selection is the process of selecting a subset of relevant features that are strongly associated with specific response variables from a high-dimensional feature set. By selecting a subset of relevant features, feature selection generally improves the efficiency of subsequent analyses, enhances the reproducibility of findings, and increases interpretability by minimizing the variables requiring further analysis

about clinical implications [41]. Consequently, this methodology is widely utilized across various research domains, particularly in studies involving preterm infants within the NICU [42-48]. Specially, many studies for developing predictive models have applied a feature selection approach for several critical objectives: mitigating model overfitting, enhancing computational efficiency, and stabilizing model performance by filtering out noisy data[42, 43, 49-51].

Broadly, feature selection methodologies are categorized into filter, wrapper, embedded, and hybrid methods [43]. In studies focused on clinical risk factor identification, the criteria for selecting or identifying risk factors were often derived from existing literature or through expert-driven clinical background. These approaches are predominantly utilized in predictive model research to extract the most relevant features from large datasets. In NICU research, the selection of a specific feature selection technique is often dictated by study characteristics. Additionally, within the clinical domain, expert-driven feature selection, which involves identifying key clinically relevant indicators, is commonly utilized [46]. For filter methods, commonly used approaches include traditional statistical techniques such as univariable logistic regression and statistical test-based feature selections [52, 53]. Specifically, stepwise feature selection approach is widely used in clinical research and practice [54, 55]. More recently, there has been an increasing adoption of embedding methods for feature selection, which utilize feature importance metrics derived from machine learning models [53, 55, 56].

Research in feature selection is rapidly advancing, driven by the increasing prevalence of data characterized by ultra-high dimensionality. Specifically, the number of false positives in feature selection procedures significantly impacts the performance of subsequent analyses and predictive models derived from high-dimensional data. Therefore, the ability to control these false positives has become

crucial, leading to recent methodological advancements in False Discovery Rate (FDR) control for feature selection methods. The FDR control method controls the proportion of false discoveries among rejected null hypotheses (i.e., irrelevant features related to specific response variables) to remain below a target level. For high-dimensional data in clinical research, the Benjamini-Hochberg (BH) procedure [57] and the Benjamini-Yekutieli (BY) procedure [58] have been widely used as FDR control methods. One notable recent development in powerful FDR control methodologies is the knockoff filter [59]. This method adds synthetic "knock-off" features into the dataset, against which the FDR is then rigorously controlled. Compared to existing FDR control methods, the Knockoff filter method is more general and flexible, providing stable FDR control even when the proportion of null features is high. Consequently, several Knockoff filter-based FDR control methods, such as Model-X, have been proposed [60, 61]. However, a significant limitation of the knock-off technique is its prerequisite for prior knowledge of the data's underlying distribution. Applying knockoff filter methods when this distribution is unknown carries the risk of FDR inflation [62]. Research is also underway to efficiently compute statistics for a large number of features. Notably, recent efforts have focused on analyzing U-statistics across distributed servers, specifically addressing the statistical-computational trade-off [63]. Despite these advancements, most NICU research predominantly utilizes classic stepwise selection methods. There's a notable lack of research and feasibility studies on feature selection methods designed to efficiently handle the large volume of features derived from continuous vital signs.

One of the primary challenges in studies aimed at identifying risk factors for preterm infants in the NICU is the limited number of subjects who meet the inclusion criteria, along with the particularly small size of the case group targeted for risk

factor analysis. This results in a highly imbalanced class distribution, which represents a key challenge in classification-based problem formulations, as outlined in the earlier methodology.

As previously noted, one of the primary challenges in studies aimed at identifying risk factors for preterm infants in the NICU is the limited number of subjects who meet the inclusion criteria, along with the particularly small. The substantial heterogeneity in demographics, institutional measurement practices, and external environmental factors among preterm infants in the NICU induced significant uncertainty into the theoretical biases of continuous vital sign-based features. Current research often attempts to identify key risk factors and predictors by ranking features based on p-values from association analyses with specific morbidities or mortality. Alternatively, basic feature selection techniques are applied for "black box" models such as deep learning and machine learning. A major limitation of these approaches is that most are single-institution studies lacking external validation, precluding confirmation of performance robustness.

1.3. Research Questions and Aims

Previous research has established that features derived from continuous vital signs can offer novel insights to clinicians and provide a foundation for identifying risk factors that enable earlier detection of preterm infant deterioration with higher sensitivity. However, current continuous vital sign processing and analysis methods vary significantly across studies. Furthermore, the analytical techniques applied are often limited by computational burden, restricting them to basic statistics, or they fail to sufficiently mitigate bias that can arise from single-institution studies.

Therefore, this study proposed a methodology designed to overcome the

limitations of existing continuous vital sign analysis and address the challenges associated with identifying novel risk factors. The overall methodological pipeline was structured into the following chapters.

Chapter 2 introduced a continuous vital sign analysis methodology that applies diverse domain methods to analyze, filter, and select features strongly associated with critical clinical events in preterm infants. This methodology uses a distributed computing architecture from its implementation phase, ensuring scalability and flexible adaptation for expanding datasets, new risk factors, and model integration in future research. Conventional FDR control and emulation methods, especially those based on traditional sequential procedures, are exceptionally time-consuming. To address this, we proposed partitionable algorithms within our framework. This transformation of iterative processes significantly enhances computational efficiency and resolves prior limitations in the scalability and flexibility of feature selection and analysis methods. By enabling parallel or distributed computing techniques, our refined approach aligns with current high-performance computing trends, thereby substantially boosting efficiency and overall scalability for larger, multicenter clinical studies.

Chapter 3 focused on identifying key features associated with all-cause mortality and late-onset sepsis, two major predictive modeling topics in NICU preterm infants. We developed predictive models and validated them using an external dataset to confirm the robustness of the methodologies proposed in this study. We also investigated whether previously known key features associated with sepsis and all-cause mortality were similarly detected, or if our approach identified complementary or surrogate features.

Chapter 4 validated the superiority of continuous vital sign-based features over existing EMR-based features in providing insight into patient clinical symptoms. To

achieve this, we recalibrated a previously developed extubation readiness model for NICU preterm infants within our study. We then confirmed the enhanced performance of this predictive model, driven by features identified in our research, using our proposed methods.

Chapter 5 demonstrated the ability of our proposed methodologies to identify novel physiological markers for the timely detection of IVH, a critical morbidity in preterm infants requiring early diagnosis and intervention. This chapter assessed the applicability of diverse cross-domain feature calculation methods in the clinical domain and explored how our approach overcomes limitations of traditional case-control studies for risk factor identification.

Chapter 2. Continuous Vital Signs Derived Feature Extraction and Analytic Frameworks for Identifying Risk Factors in Preterm Infants

2.1. Introduction

In recent NICU studies, continuous vital signs, sampled at higher frequencies (0.03 Hz to 2 Hz) than traditional EMR data, are being analyzed for their association with major complications and mortality in preterm infants, and for developing predictive models [34, 35, 39, 64]. These predictive models demonstrated superior performance compared to EMR-based approaches, and comparable accuracy to ECG-based predictive models, implying significant clinical potential. These data are being used not only for monitoring but also for identifying physiological markers, predictors, and risk factors that are not detectable using conventional EMR-based data.

Even with the significant potential of continuous vital signs, the considerable computational demands remain a major challenge to advanced research. For this reason, many studies identifying clinically relevant variables are still limited to extracting continuous vital sign features based on basic descriptive statistics or established measures, such as HRC and HRV. The discovery, development, and validation of novel time series analysis methods inherently necessitate significant time and computational resources. Consequently, researchers in medical engineering, deep learning, and medical statistics dedicate substantial effort to validating newly identified or proposed clinical risk indicators and assessing their applicability across diverse clinical contexts. Notably, unlike adults, analyzing continuous vital signs in

preterm infants requires specific methodological consideration due to the variation in their physiological characteristics with developmental maturity. Furthermore, analytical challenges are compounded by small patient cohorts, the low incidence of major comorbidities and mortality, and the frequently unidentifiable precise onset times of adverse clinical events. Therefore, research on analytic methodologies and frameworks is essential to accelerate early-stage continuous vital sign studies and mitigate the aforementioned challenges. However, current frameworks and analytical methodologies within the clinical domain lack the maturity to provide the comprehensive understanding required by all stakeholders.

To address this, some studies have applied time-series feature extraction techniques, either domain-specific or domain-agnostic, to derive more informative representations of NICU monitoring data [34, 35, 39]. These methods aim to retain the temporal complexity of the original signals and improve clinical relevance. However, most studies to date have only confirmed previously known features rather than identifying new or context-specific ones [34]. Additionally, downsampling techniques, such as random selection or averaging (e.g., grand mean), were applied to manage the high sampling frequency [30, 34, 35, 39]. While these approaches simplify processing, they might result in a loss of temporal resolution and fail to capture transient physiological changes. Moreover, the feature selection and analysis methods applied in the clinical studies have generally lacked scalability, limiting their applicability in larger, multicenter studies or real-time clinical settings.

This chapter introduces a methodological framework for extracting and efficiently selecting novel physiological features that are linked to major complications and mortality in preterm infants. Unlike previous studies that commonly applied downsampling techniques, this approach supports the full temporal resolution of the data to capture subtle physiological dynamics that may be

clinically significant. It is specifically designed to support real-time feature extraction and analysis in routine clinical settings. To assess its computational efficiency, the proposed method was evaluated against established feature selection algorithms by comparing processing time and resource usage under equivalent feature sets. In addition, the characteristics of the features generated through the proposed method will be examined in relation to specific complications and mortality outcomes.

2.2. Methods

2.2.1. Design Principles

This study focused on designing and implementing continuous vital signs analytical methodologies and frameworks by systematically addressing the identified limitations of existing continuous vital signs analysis. Furthermore, we proposed algorithms to address analytical challenges resulting from the heterogeneous nature of NICU patient populations across countries and institutions, including variations in intervention procedures, demographics, and incidence rates of mortality and morbidity.

To mitigate the insufficient accessibility of existing time series analysis methods, other valuable analytical techniques, and domain-agnostic feature calculations, we applied “off-the-shelf” approaches. In other words, the proposed framework addresses the substantial computational demands in clinical domain research, particularly those arising from the high-order time complexity associated with features combining multiple time series segments (e.g., sample entropy), by precomputing all necessary features via seamless libraries integration. These

libraries consist of either directly implemented or externally imported essential feature calculations. It also offers the flexibility to integrate methods from other domains or novel AI models into the clinical setting.

Continuous vital sign data often yield high-dimensional feature sets, which pose significant analytical challenges. Specifically, the multiple comparisons inherent in feature selection processes can lead to a very high rate of false positives. Furthermore, many feature selection methods require a large number of iterations to converge, adding to computational burden. Moreover, analysis is complicated by immaturity effects observed in preterm infant populations and avoidable biases frequently encountered in observational studies. These biases include immortal time bias, depletion of susceptible bias, confounding, and the false discovery problem. To address these issues, we implemented several advanced techniques. First, we refined and applied state-of-the-art FDR control methods to manage false positives effectively. Second, to account for biases common in observational studies, we applied an emulation of a matched case-control design.

Crucially, implementing both FDR control methods and emulation, particularly when based on traditional sequential procedures, can be exceptionally time-consuming. Therefore, we proposed algorithms that refined these two methods by transforming traditional iterative algorithms into partitionable algorithms within our frameworks, significantly enhancing computational efficiency. This refined approach addresses the challenge of limited scalability and flexibility in feature selection and analysis methods, which has historically restricted their applicability in larger, multicenter clinical studies. Moreover, by enabling the utilization of parallel or distributed computing techniques, which align with current trends in high-performance computing, it substantially enhances computational efficiency and overall scalability.

2.2.2. Time Series Analysis-based Feature Extraction

Approach

Continuous vital sign data, periodically recorded from patient monitors, consists of high-frequency, high-resolution numeric measurements of multiple physiological parameters, including heart rate, pulse, oxygen saturation, invasive arterial pressure, and respiratory rate. These inherent characteristics allow continuous vital sign data to be considered a distinct form of time series data. These signals exhibit substantial temporal complexity and high dimensionality, making direct interpretation and modeling challenging.

Given the inherent time series characteristics of continuous vital sign data, we applied prominent time series analysis methodologies, such as time-domain, frequency-domain, linear correlation, and information theory approaches, for feature extraction [65-68]. Time series feature-based analysis is a widely used approach across most domains, and as a result, extensive research and various implementations for feature extraction have already been developed [66]. These methods were selected to effectively capture the temporal structure and complexity of high-resolution physiological time series data, beyond the capabilities of conventional analytical approaches to extract clinically meaningful patterns and information that may aid early risk assessment and support timely intervention in neonatal care.

2.2.3. Continuous Vital Signs Feature Calculation Methods

Due to the need to identify subtle, previously unrecognized physiological risk factors and predictors of adverse outcomes in preterm infants, this study applied a wide range of feature extraction and extraction methods drawn from diverse domains. To capture the diverse temporal and statistical characteristics of high-resolution

physiological signals, we applied a wide range of time series feature extraction techniques grounded in well-established mathematical and statistical principles. These features were categorized into the following theoretical classes to ensure both interpretability and comprehensive signal representation.

Descriptive statistical methods were applied to characterize the overall distributional properties of the continuous vital signs. These included measures of central tendency, dispersion, and higher-order moments (e.g., skewness and kurtosis), which capture asymmetrical and tail behavior [69, 70]. These features provide essential baselines for identifying abnormal clinical symptoms, such as desaturations which may easily be identified by descriptive statistics [69, 71]. These features are widely adopted in studies utilizing continuous vital signs, serving as a comprehensive framework for physiological data analysis [30, 35, 72].

Time-domain analysis methods were implemented to capture local dynamics and signal shape characteristics [65, 73]. Features such as the longest consecutive runs above or below the mean, first and last positions of local maxima and minima, and threshold-based counts were used to quantify signal excursions, volatility, and event durations [66, 74]. These properties are particularly relevant in detecting transient physiological events, such as apnea episodes or episodic desaturation, which may not be evident in aggregated summary statistics.

We also implemented correlation-based features to model temporal dependence and repetitive patterns within the signals [67, 68]. Classical autocorrelation and partial autocorrelation functions were used to identify short- and long-range dependencies, providing insight into regulatory patterns in cardiorespiratory signals [36].

Frequency-domain and spectral methods were computed to analyze the distribution of power across different frequency bands [67, 68, 75-79]. Fourier

analysis and wavelet-based methods were applied to extract both stationary and transient periodic components, with the aim of capturing lower-frequency characteristics not typically observed in conventional ECG frequency features [13, 14]. Welch's method [80] was used to estimate power spectral density with reduced variance.

Entropy and information theory-based methods were also applied to quantify the regularity, complexity, and unpredictability of the continuous vital sign time series [81-83]. Features derived from approximate entropy, sample entropy, permutation entropy, and Lempel-Ziv complexity[84] provide estimates of signal irregularity based on symbolic representations or probability distributions. These metrics are particularly sensitive to subtle changes in physiological regulation and have been associated with pathophysiological states such as sepsis, neurological instability, and poor autonomic tone [30, 81].

Lastly, Linear regression- and model-based methods were derived by fitting linear models and stochastic differential equation (SDE) approximations to segments of the vital sign data [66, 74]. Linear trend coefficients describe the direction and rate of physiological change over time, while Friedrich coefficients capture drift and diffusion characteristics under the assumption of nonlinear stochastic dynamics. These features allow for interpretable modeling of trends, such as progressive bradycardia or deteriorating oxygenation, which unfold over longer durations.

Additionally, considering that vital sign data from patient monitors are recorded as integers, the index of qualitative variation (IQV) [85-91] was also applied to account for this characteristic. Vital sign data collected from patient monitors, including heart rate, respiratory rate, and peripheral oxygen saturation, are typically stored as discrete integer values due to limitations in hardware precision and the requirements of real-time clinical monitoring. Unlike continuous, high-resolution

waveforms such as ECG, these discretized signals offer limited granularity, which can obscure subtle physiological variations when analyzed using conventional statistical measures. To address this issue, the IQV was incorporated into the feature extraction process. IQV is particularly effective for analyzing non-continuous or discretized data, as it captures distributional variability that traditional metrics like mean and standard deviation may overlook, especially when the data have been rounded or encoded at low resolution. By applying IQV, the feature set gains additional representational depth, which can be especially beneficial in studies with small sample sizes or imbalanced class distributions. In such cases, reliance solely on standard numerical descriptors can compromise model performance, whereas the inclusion of qualitative variability measures offers a complementary approach to detecting clinically relevant patterns.

We utilized several Python libraries, including NumPy [92], SciPy [93], tsfresh [74], statsmodels [94], and librosa [95], to implement the previously described feature extraction techniques. For algorithms not provided by these packages, we implemented the required methods. To optimize performance and mitigate the performance degradation resulting from Python’s iteration process, we compiled time-critical code sections with LLVM using Numba [96].

2.2.4. Continuous Vital Signs Feature Analysis and Selection

We defined the analytical problem as a multivariate time series classification task to optimize the interpretation and utility of the analysis results. Based on this problem definition, we designed an effective methodological framework to conduct statistical analysis and hypothesis-driven feature evaluation aimed at identifying variables significantly associated with adverse clinical outcomes such as complications and

mortality. To achieve this, we utilized various domain-agnostic feature selection algorithms [74], which were combined or refined as necessary, to meet the aims of this study as schematically represented in Figure 2-1.

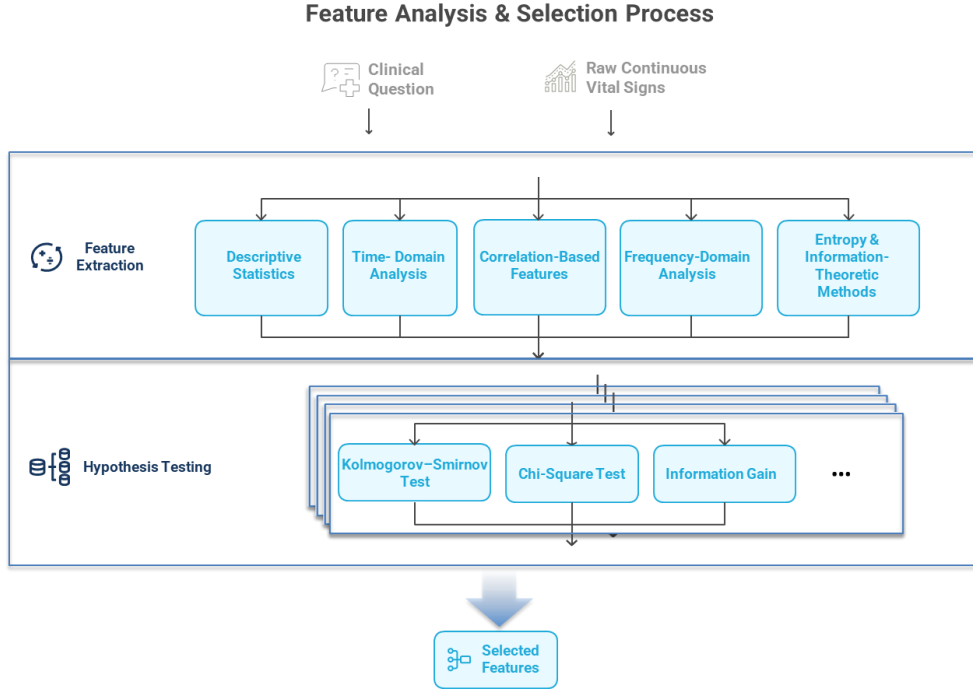


Figure 2-1. Proposed feature analysis and selection method.

Features derived from continuous vital signs present several distinct analytical challenges. Firstly, the large number of extracted features and their unclear physiological implications challenge traditional clinical expert-driven selection approaches. Secondly, this dataset inherently exhibits high dimensionality, where the number of features p frequently exceeds the available sample size n . Lastly, the dynamic growth trajectory of preterm infants leads to significant time-dependent variations in vital sign-based features.

Research in preterm infants, characterized by a large number of candidate predictors and risk factors relative to a small, highly heterogeneous patient population, shares methodological similarities with Genome-Wide Association

Studies (GWAS). In single nucleotide polymorphism (SNP) studies, bootstrapping and FDR control methods have been frequently used to identify robust risk factors and associations [97-99]. Therefore, this study proposed methods for efficiently controlling the FDR in large feature sets while enabling parallel computing and aims to implement these methods as a framework.

This study implemented a multiple data-splitting FDR procedure, adapted from the method of Dai, et al. [41] for distributed computing. Dai, et al. originally proposed a data-splitting approach to FDR control, based on Least absolute shrinkage and selection operator (Lasso) and Ordinary Least Squares (OLS) regression, to overcome the limitations of conventional FDR methods. Their framework demonstrated high efficiency and strong performance in relevant feature selection within high-dimensional datasets.

We adopted this FDR control framework for several reasons. Firstly, data splitting, which involves randomly partitioning a dataset for analysis, inherently facilitates distributed processing. Secondly, the applicability is straightforward, requiring only that the estimated coefficients be symmetric around zero, a condition easily met in our context.

However, a notable drawback of the original study is its dependence on Lasso regression to derive coefficient values. As the feature dimension increased, the number of iterations required for convergence, and consequently the computational time, substantially escalates. Given the nature of convex optimization problems, Lasso-based feature selection methods are substantially limited in $p > n$ scenarios [100]. Furthermore, the original Lasso implementation is restricted in the parallelization transformation because its convergence requires the outputs from prior iterations. Therefore, to retain the advantages of FDR control via data splitting while enabling effective parallel processing, we modified the original approach for application in this research.

In this study, we formulated mirror statistics based on the equations [41, 101].

Mirror statistics exhibit several key properties. First, relevant features, which those highly associated with response variables or events, yield larger positive mirror statistics. Second, irrelevant or null features tend to produce mirror statistics that are close to zero or symmetric around zero.

These characteristics are achieved by merging the coefficients from two split estimators. This design offers a significant advantage. Even if a false positive association inadvertently arises from one estimator, the differing association from the other estimator helps ensure the merged result is symmetric around zero. This makes the mirror statistics more robust than those derived from a single estimator and facilitates straightforward application. We used the following equation to calculate mirror statistics M_j in this study.

$$M_j = \text{sign}(\hat{\beta}_j^{(1)} \hat{\beta}_j^{(2)}) f(|\hat{\beta}_j^{(1)}| |\hat{\beta}_j^{(2)}|),$$

where $\hat{\beta}_j^{(1)}, \hat{\beta}_j^{(2)}$ are the estimated coefficients for the features obtained from each data split, and merging function $f(u, v) = uv$. To compute the mirror statistics, we utilized estimators based on the Chi-Squared test, information gain[102], and the Kolmogorov-Smirnov(KS) test [103]. For the Chi-Squared test, numerical variables were binned into a 2×2 contingency table using the chi-merge method [104]. To estimate $\hat{\beta}^{(1)}$, numerical variables were transformed into 2×2 contingency tables using the chi-merge method. The resulting Chi-Squared test p-values were then adjusted using the FDR-BY procedure and subsequently merged with the signs derived from OLS regression coefficients. This merging with the OLS-derived sign was performed to ensure the resulting statistics were symmetric around zero. $\hat{\beta}^{(2)}$ was calculated based on information gain statistics and subsequently combined with results from the KS test to achieve symmetry around zero. In Dai, et al. [41] study, the function $f(u, v) = u + v$ was chosen as the optimal transformation to ensure

the sampling distribution of the mirror statistics M_j maintained symmetry around zero and simultaneously achieved maximal statistical power. However, in the context of this study, where the Chi-Squared test is adopted and a 2×2 contingency table arrangement results in Chi-Squared statistics equivalent to squared Z-test statistics, we opted for $f(u, v) = uv$ to fulfill the crucial symmetry assumption for the mirror statistics. The cutoff methodology in this study follows that of the data-driven cutoff approach proposed by Dai, et al. [41]. The data-driven cutoff τ_q as followed.

$$\tau_q = \min \left\{ t > 0 : \frac{\#\{j: M_j < -t\}}{\#\{j: M_j > t\}V1} \leq q \right\},$$

where $q \in (0,1)$ is the target FDR control level, and M_j is the mirror statistics of the j th feature.

Feature selection for each subsample in this research was subsequently conducted by implementing Algorithm 1.

Algorithm 1. Feature Selection via false discovery rate control with multiple data split replications

Input: $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$: original dataset

m : number of data split replications

q : FDR control level

Output: Selected relevant feature set $\hat{S}$

1. Initialize $S \leftarrow \emptyset$

for $i = 1$ to m **do**

2. Split the data into two groups $D^{(1)}, D^{(2)}$.

3. Estimate the coefficient $\hat{\beta}^{(1)}$ from Chi-squared statistics (chi-merge) with OLS using $D^{(1)}$.

4. Estimate the coefficient $\hat{\beta}^{(2)}$ from information gain and Kolmogorov-Smirnov test using $D^{(2)}$.

5. Calculate the mirror statistics $M_j = \text{sign}(\hat{\beta}_j^{(1)} \hat{\beta}_j^{(2)}) f(|\hat{\beta}_j^{(1)}| |\hat{\beta}_j^{(2)}|)$.

6. Calculate the cutoff value $\tau_q = \min \left\{ t > 0: \frac{\#\{j: M_j < -t\}}{\#\{j: M_j > t\} \vee 1} \leq q \right\}$.

7. Select the features $S^{(i)} \leftarrow \{j: M_j > \tau_q\}$ where τ_q is the cutoff for FDR level q .

8. Append $S^{(i)}$ to S .

for $j \in \{1, 2, \dots, p\}$ **do**

9. Estimate the associated inclusion rate $\hat{I}_j = \frac{1}{m} \sum_{k=1}^m \frac{\mathbb{I}(j \in S^{(k)})}{|S^{(k)}| \vee 1}$.

10. find the largest $\ell \in \{1, 2, \dots, p\}$ that satisfies $\hat{I}_1 + \hat{I}_2 + \dots + \hat{I}_\ell \leq q$.

11. select the features $\hat{S} = \{j: \hat{I}_j > \hat{I}_\ell\}$.

Consequently, proposed methods were specifically designed to address two methodological considerations. The first consideration is the crucial issue of FDR control within a high-dimensional feature space. The second involves strategically mitigating the confounding effects of immaturity and ongoing growth in preterm infants. As mentioned in Chapter 1, preterm infants showed substantial inter-individual variability in clinical severity directly result from their developmental immaturity. As these infants mature, they progressively acquire the essential physiological functions required for extrauterine survival. These developmental changes lead to considerable fluctuations in vital signs. A critical challenge arises because major clinical events, routinely investigated in NICU research, often

manifest at disparate maturational stages, thereby necessitating robust analytical strategies to account for these developmental discrepancies. Fundamentally, this work endeavors to minimize avoidable biases induced from observational studies, including immortal time bias, depletion of susceptible bias, confounding, and the false discovery problem [105].

To mitigate the effects of immaturity observed in our observational study, this research employed an emulation of a matched case-control design. This approach aligns with a growing trend in recent clinical research and guidelines, where the emulation of a target trial is increasingly applied to address the inherent limitations of conventional observational studies [106]. Applying this target trial emulation not only mitigates avoidable biases inherent in observational studies but also offers the significant advantage of reducing ethical concerns and potential harm to patients [105-107]. Therefore, leveraging these advantages, this study implemented an emulation of a matched case-control study through bootstrapping, based on the method outlined below. This approach was specifically motivated by research focused on emulating patient target trials [108]. The emulation of a matched case-control design was implemented as a procedure that aggregates feature selections performed within subsamples, following Algorithm 1 as previously described. We denoted the dataset as $\mathcal{D} = \{(x_i, y_i, d_i)\}_{i=1}^N$, where $x_i \in \mathbb{R}^p$ is the continuous vital signs feature vector, y_i is the response variables, and d_i is the demographics information of preterm infants i .

Algorithm 2. Feature selection and aggregation via emulations**Input:** $\mathcal{D} = \{(x_i, y_i, d_i)\}_{i=1}^N$: original dataset m : number of resamples $c \in [0,1]$: inclusion rate cut-off threshold $\mathcal{J}$: function for inclusion criteria $\mathcal{F}(\cdot)$: the proposed feature selection methods (Algorithm 1.)**Output:** Selected relevant feature set $\hat{S}$ 1. Initialize $S \leftarrow \emptyset$ **for** $i = 1$ to m **do**2. Generate a subsample $D^{(i)}$ of size N from $\mathcal{D}$,
restricted to sample satisfying $\mathcal{J}(d_i) = \text{True}$, with replacement.3. Let $X^{(i)} \in \mathbb{R}^{n_i \times p}, Y \in \mathbb{R}^{n_i}$
denote the feature and response vector corresponding to $D^{(i)}$.4. Apply the feature selection methods, $S^{(i)} \leftarrow \mathcal{F}(X^{(i)}, Y^{(i)})$.5. Append $S^{(i)}$ to S .**for** $j \in \{1, 2, \dots, p\}$ **do**6. Estimate the associated inclusion rate $\hat{I}_j = \frac{1}{m} \sum_{k=1}^m \frac{\mathbb{I}(j \in \hat{S}^{(k)})}{|\hat{S}^{(k)}|_{V1}}$.7. find the largest $\ell \in \{1, 2, \dots, p\}$ that satisfies $\hat{I}_1 + \hat{I}_2 + \dots + \hat{I}_\ell \leq c$.8. select the features $\hat{S} = \{j: \hat{I}_j > \hat{I}_\ell\}$.

Finally, to enable the application of the aforementioned feature selection methods and the bootstrap-based subsample feature selection and aggregation procedure to parallel computing, we configured the system as follows (Algorithm 3). Our proposed algorithm enhances efficiency by not copying and transferring the entire dataset. Instead, it provides only the keys $\mathcal{T}$ required to access the data, enabling efficient task set partitioning and registration to individual server job pools. Additionally, this design inherently simplifies partition configuration, as each unit task of $\mathcal{T}$ solely requires ensuring result collection at its terminal point. Although Algorithm 3 defines K in terms of server units, it offers the flexibility to partition K based on alternative criteria as necessitated by varying contexts. Beyond this, the system can be structured to support dynamic resource allocation and processing by integrating load balancing capabilities.

Algorithm 3. Parallelized replication for feature selection in high-dimensional data

Input: $\mathcal{D} = \{D^{(i)}\}_{i=1}^m$: subsampled datasets
 $D^{(i)} = \{(x_j^{(i)}, y_j^{(i)})\}_{j=1}^{n_i}$: the i th subsample dataset

m : number of resamples

p : number of features

K : number of servers

$\mathcal{F}(\cdot)$: the proposed feature selection methods (Algorithm 1.)

Output: Selected relevant feature set $\hat{S}$

Part I: Construct job list

1. Construct task set of all Mp jobs:

$$\mathcal{T} = \{(i, j) | i \in 1, 2, \dots, m; j \in 1, 2, \dots, p\}$$

2. Flatten the task set, and divide into K disjoint parts:

$$\mathcal{T}' = \cup_{k=1}^K \mathcal{T}^{(k)}, \text{ where } \mathcal{T}^{(k)} = \{(i, j) \in \mathcal{T} | \text{flat index} \in [\frac{kMp}{K} + 1, \frac{(k+1)Mp}{K}]\}$$

Part II: Distribute to servers

3. Send each task set $\mathcal{T}^{(k)}$ to k th server.

Part III: Local computation

4. On the k th server, compute feature selection methods

$$S_{i,j} = \mathcal{F}(x_j^{(i)}, y_j^{(i)}), \text{ where } (i, j) \in \mathcal{T}^{(k)}$$

Part IV: Aggregation

5. Gather all computed selected feature sets from servers, reconstruct matrix:

$$S = [S_{i,j}] \in \mathbb{R}^{m \times p}$$

for $j \in \{1, 2, \dots, p\}$ **do**

6. Estimate the associated inclusion rate $\hat{I}_j = \frac{1}{m} \sum_{k=1}^m \frac{\mathbb{I}(j \in \hat{S}^{(k)})}{|\hat{S}^{(k)}|_{V1}}$.

7. find the largest $\ell \in \{1, 2, \dots, p\}$ that satisfies $\hat{I}_1 + \hat{I}_2 + \dots + \hat{I}_\ell \leq c$.

8. select the features $\hat{S} = \{j: \hat{I}_j > \hat{I}_\ell\}$.

2.2.5. Data Analysis Framework Implementation

In the implementation phase, we evaluated the scalability and parallel processing capabilities of these computationally demanding methods, as detailed previously. Therefore, this study proposed and implemented the scalable continuous vital sign analysis process, assuming the infrastructure outlined in the scenario is already in

place (Figure 2-2).

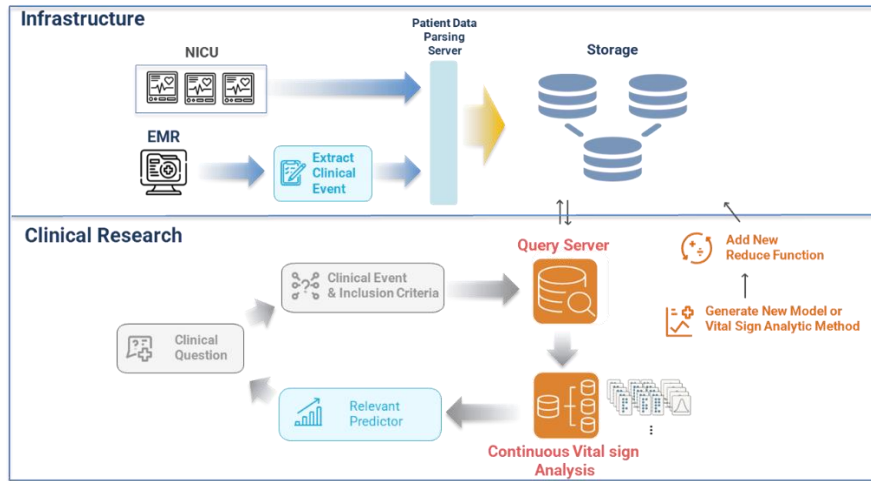


Figure 2-2. Proposed continuous vital signs analysis process.

We considered the following design requirement. First, the framework has to simultaneously process large volumes of continuous vital sign data. Unlike EMR data, which are typically recorded selectively based on clinical judgment or intervention, continuous vital sign data are collected automatically and non-selectively through patient monitors in real time, without being influenced by clinician intent. This objective and uninterrupted acquisition results in datasets that are substantially larger relative to the sampling rate, often requiring computational strategies designed for large-scale time-series data. To address the resulting analytical demands, the MapReduce computing model (Figure 2-3) was applied within the proposed framework to effectively handle the volume and structure of the data [109, 110].

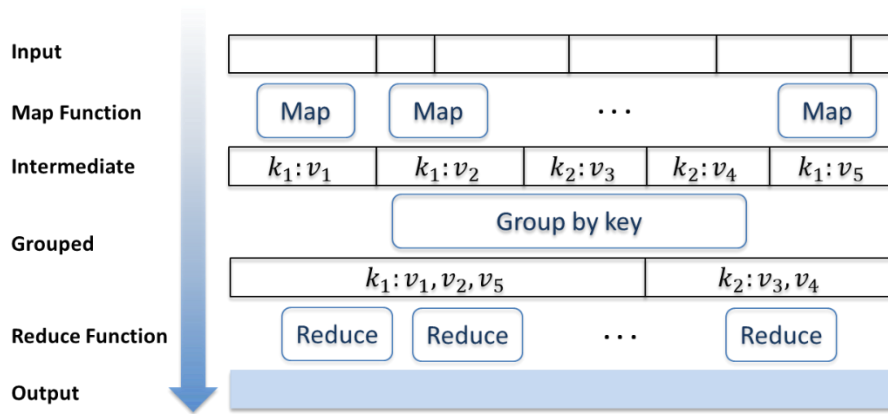


Figure 2-3. MapReduce model [109].

While a MapReduce model is not optimized for real-time analysis or typical healthcare IT environments, we adopted it for this study due to several key advantages. First, it offers an intuitive and immediately applicable framework for scalable continuous vital sign analysis. Second, the MapReduce paradigm is a proven system for big data analysis, validated across numerous fields. Finally, its inherent Split-Apply-Combine strategy facilitates the extensibility of bootstrapping methods, which was crucial for this study [111]. By directly implementing this existing distributed computing paradigm, we aimed to demonstrate the compatibility of our proposed continuous vital sign feature extraction and analysis methods with modern distributed computing architectures.

To assess the compatibility of our proposed methods with the fundamental MapReduce paradigm, this study utilized CouchDB [112]. CouchDB is a document-based NoSQL database specifically designed with the MapReduce computing model as a core architectural principle.

Most NoSQL databases such as MongoDB, Aerospike, DynamoDB, Azure Cosmos DB, Apache Ignite, and Cassandra, which often rely on separate aggregation functions or in-memory approaches for rapid response, leading to performance

variations based on optimization and server computational resources [113-120]. In contrast, CouchDB's document store architecture provides a stable and predictable environment. This made it particularly suitable for evaluating the minimum performance baseline of the methodologies we proposed.

Furthermore, CouchDB aligns well with the practical needs of a typical clinical institution. It facilitates incremental server expansion and contraction, which is crucial for adapting to fluctuating demands. More importantly, for use in a clinical setting, it offers high fault tolerance and availability, making it an ideal choice for developing a robust database for this study.

In MapReduce model, implemented through CouchDB's view system, we defined the three distinct views to process the input data and generate defined document outputs. These views—an event view, a feature view, and a demographics view—were designed to efficiently extract features from original vital signs, define clinical events, and generate random subsamples (Figure 2-4).

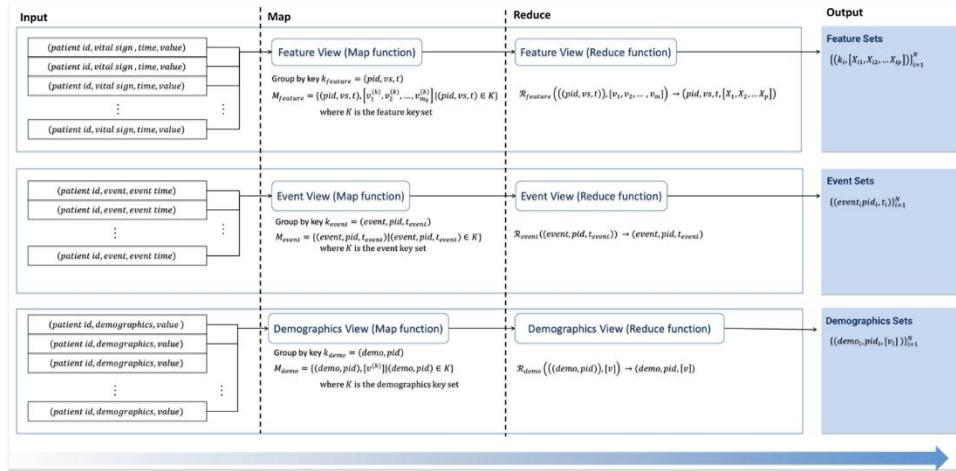


Figure 2-4. Map-reduce workflow for subsampling and feature extraction.

The Event view was specifically configured to capture the event timelines of the preterm infants under investigation. Within its Map function, patient identifier,

event name, and event time were defined as the group key. This view processed emitted data as individual documents and did not implemented any reduce operations.

This design choice was made because, during the emulation of the analysis enrollment process, only the event and demographic information for the study cohort is required. Should additional document-specific details be necessary, it is more efficient to query these documents subsequent to the enrollment phase; thus, a Reduce function was not applied within this view.

The Demographics view was designed to capture the demographic information of the preterm infants. In its Map function, patient identifier and key demographic variables, such as gestational age and birth weight, were configured as the group key. This view, similar to the Event view, did not define any Reduce function. This design choice was based on the efficiency gained by performing initial patient screening at the enrollment stage, with more detailed document queries executed as needed post-enrollment.

The Feature view processed original continuous vital sign time series data. Its Map function rounded measurement timestamps to the nearest hour, defining the patient identifier, specific vital sign, and rounded measurement time as the key. This data then passed through a Reduce function, generating the diverse domain-specific features discussed in the preceding section. While CouchDB's views typically use JavaScript Map-Reduce, we implemented distributed processing by loading these operations to a separately configured Docker-based Python query server container. This design was specifically chosen to enhance processing speed and improve the extensibility of feature calculation functions.

The implementation of the case-control study emulation and the feature selection methods followed the workflow depicted in Figure 2-5.

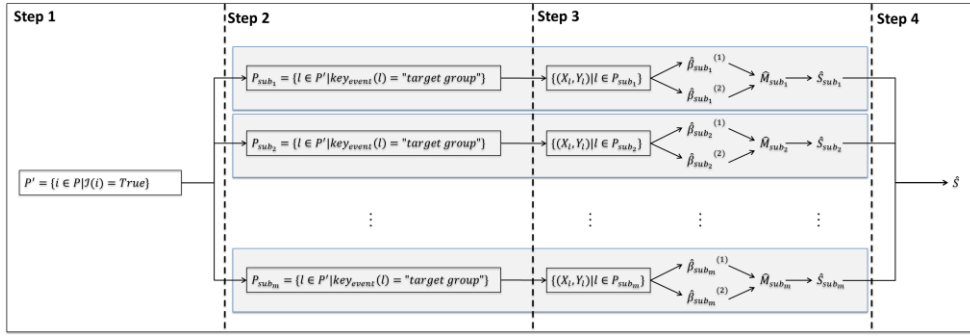


Figure 2-5. Integrated workflow for emulating Case-Control studies and feature selection.

The feature selection workflow, implemented in this study, comprises four distinct steps based on Algorithm 3, as introduced in the preceding steps.

Step 1: This initial step involves extracting patient identifiers that satisfy the predefined inclusion criteria for the infants under investigation. Specifically, patient identifiers for both the event and control groups are first retrieved from the Event view. Subsequently, corresponding demographic entries and patient identifiers matching these inclusion criteria are extracted from the Demographics view for verification.

Step 2: This step performs the actual bootstrapped emulation of a case-control study. Patient identifiers are randomly selected based on numbers generated by a random generator to form subsamples. From these subsamples, feature sets for analysis are extracted using patient identifiers, event time, and vital signs as keys.

Step 3: In this step, relevant features are selected utilizing the methods and procedures implemented in Algorithm 1 of this study.

Step 4: The final step involves re-selecting relevant features from each subsample based on their inclusion rate, as detailed in Algorithm 2 and 3, to yield the set of relevant features.

This implementation, structured as described, offers significant parallel

processing capabilities. With the exception of Step 4 (the aggregation phase), all tasks were designed as independently computable units. This flattened task set allows for several levels of parallelization: local parallel processing of feature estimation within individual subsample nodes, across different features, or even full parallel processing of feature estimators based on subsample index.

Consequently, if additional computational nodes or parallel processing resources are available, the system can flexibly handle multiple subsamples and their corresponding feature coefficient estimations in a distributed manner prior to the aggregation step.

A key consideration for the overall completion of this algorithm is that its subsampling procedures cannot terminate until the full computation of all feature estimators (Steps 2–3) has been performed for relevant feature selection. This scheduling consideration is a common and critical factor not only for this framework but also for many parallel computing tasks. Fortunately, various methodologies and solutions have been proposed to mitigate and resolve such issues, and these can be similarly applied to the proposed framework in this study.

2.2.6. Validation Strategies

To validate the hypotheses underlying the methodologies and the implemented framework proposed in this study, we first generated a synthetic dataset and conducted a simulation study. The primary objective of this simulation study was to determine whether our parallel approaches effectively reduce the computational burden, particularly computation time, compared to conventional methods. Concurrently, we aimed to verify the robustness of our methods when confronted with features of unknown distributions. We also evaluated to characterize and validate the performance of the proposed methods by evaluating the FDR and feature

selection sensitivity for relevant features, considering varying levels of signal power, sample sizes, and ratios of null (irrelevant) features.

Furthermore, we focused on validating the proposed methods using real-world data pertinent to key NICU research topics. Simultaneously, this study validated the proposed methods by identifying clinically significant variables and evaluating their capacity to substantially enhance predictive model performance. Our initial objective was to explore the generalizability and robust performance of our predictive models. This included evaluating their ability to identify clinically critical event predictors and also assessing the performance of the predictive models within an external validation cohort. We also validated whether the identified features were consistent with or distinct from previously established characteristics, thereby evaluating our feature selection aligned with intended clinical interpretations. This allowed us to determine the features were selected as intended. Then, we compared the performance of an existing extubation readiness prediction model, which relied on descriptive statistics. This was contrasted with a model incorporating features extracted using our proposed continuous vital sign analysis methodology. This comparison assessed whether our approach yielded superior performance compared to existing descriptive statistics. Lastly, to determine if our methodology could identify previously unrecognized physiological predictors or risk factors, we extracted vital sign risk factor identification markers for IVH using the proposed framework.

2.3. Results

2.3.1. Evaluation of Parallel Procedure in Execution Time

We evaluated the computational benefits of the proposed parallel procedure. We

conducted simulation study to compare the execution time of our parallel approach against a traditional sequential procedure across a range of data complexities, from low-dimensional to high-dimensional datasets. We combined sample sizes n of 64, 256, 512, and 2048 with feature dimensions p of 10, 100, 1,000, and 10,000.

For the experimental setup, a single Kubernetes [121] cluster, configured with Kind [122], was deployed on the single server. This cluster consisted of three nodes, each allocated 8 cores for this analysis. Mean and standard deviation (SD) of execution times were calculated from 200 repetitions. The execution time results are presented in Figure 2-7, Figure 2-8.

For the feature dimension of $p=10$, the sequential procedure demonstrated a marginally faster mean (SD) execution time of 0.376 (0.122) seconds compared with the parallel procedure's 0.583 (0.139) seconds. However, the computational advantages of the parallel approach became apparent as the number of features increased. At $p=100$, the parallel procedure completed in 1.232 (0.367) seconds, which was significantly faster than the sequential procedure execution time 2.481 (0.569) seconds. This performance divergence was increasingly pronounced with higher feature dimensions. For $p=1,000$, the parallel procedure completed in 2.574 (0.507) seconds, whereas the sequential procedure required 22.016 (4.225) seconds. At $p=10,000$, the parallel procedure demonstrated significantly faster execution at 15.060 (2.293) seconds, compared to 221.830 (59.631) seconds for the sequential procedure.

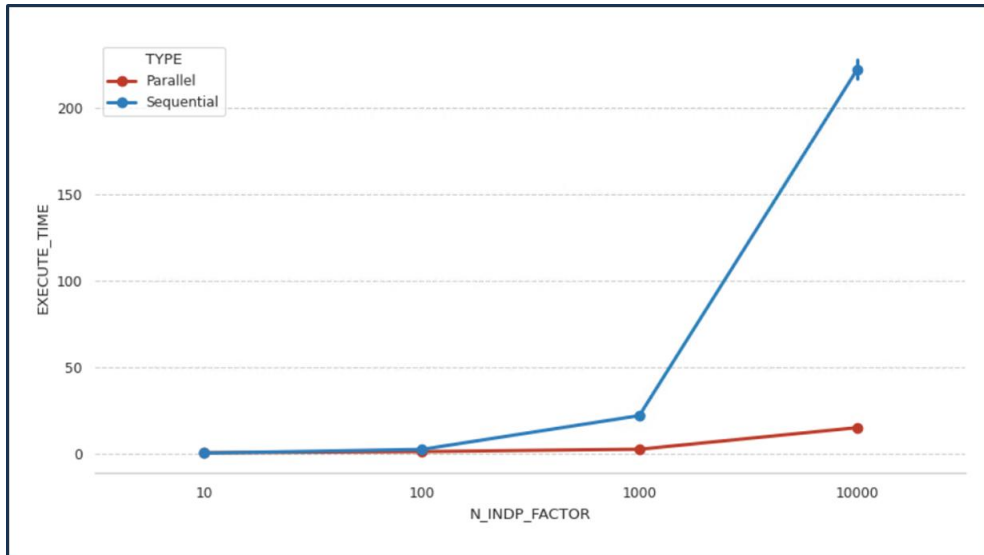


Figure 2-6. Comparison of execution times for parallel and sequential procedures across varying feature dimensions.

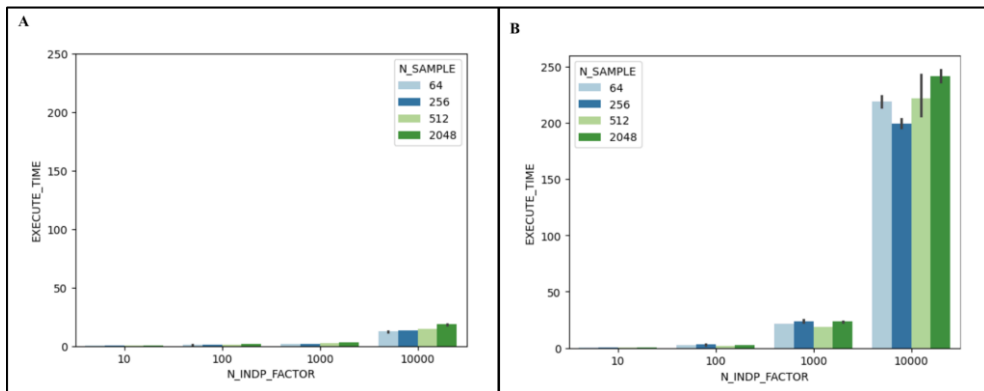


Figure 2-7. Execution time by sample size and the number of features: (A) parallel procedure, (B) sequential procedure

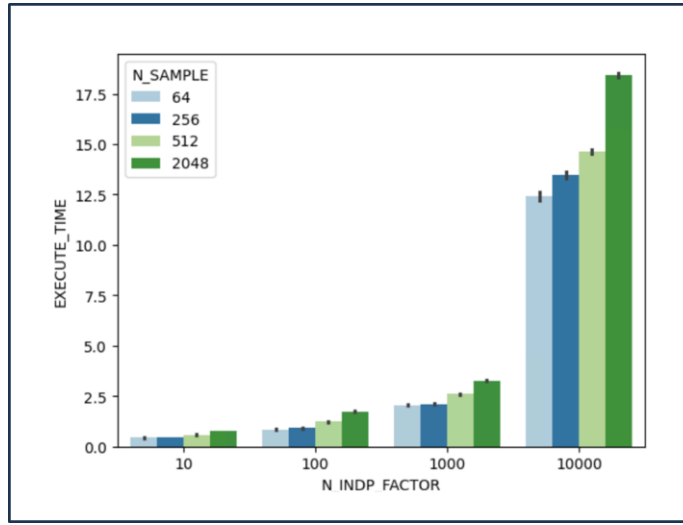


Figure 2-8. Execution time by sample size and the number of features in parallel procedure.

2.3.2. Evaluation of FDR and Sensitivity for relevant features

To evaluate the performance of our proposed feature selection methodologies, we conducted a simulation study based on the regression problem formula from Frieman's [123] study, as followed:

$$f(X) = 10 \sin(\pi x_1 x_2) + 20 \left(x_3 - \frac{1}{2}\right)^2 + 10x_4 + 5x_5 + \epsilon.$$

This formula allowed us to assess the sensitivity to relevant features and the FDR. For this analysis, we set the sample size at 256 and fixed the number of true relevant features at 5 and null features at 10,000. We then varied the case ratio at 0.1, 0.25, and 0.5 for different simulation runs. These ratios were selected to reflect typical case-control patient ratio and feature dimensions commonly observed in NICU research.

The results of the simulation runs are presented in Figure 2-9 and Figure 2-10. Our proposed methodologies consistently demonstrated high sensitivity, even at low

event rates. Notably, while traditional logistic regression-based feature selection yielded a comparable FDR to our methods, it was unable to identify relevant features. Conversely, the feature selection based on KS test and Chi-Squared test showed high sensitivity at the event ratio of 0.1, but their FDR approached 1.0. This result indicates a substantial number of null features were incorrectly identified as false positives.

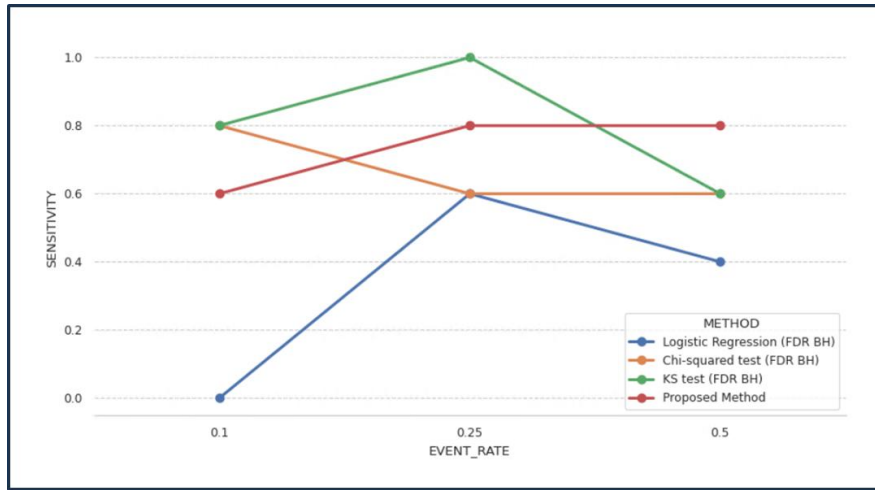


Figure 2-9. Sensitivity results from the simulation study.

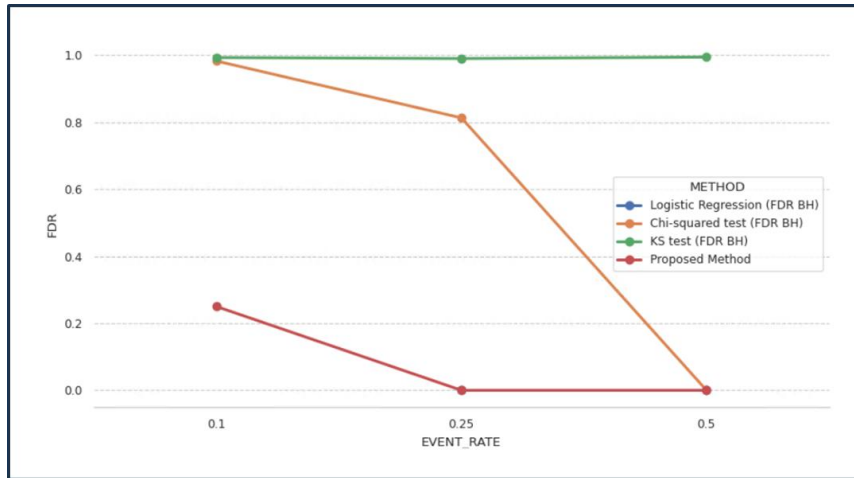


Figure 2-10. False discovery rate results from the simulation study.

Therefore, our methodologies exhibited stable feature selection performance even with a low case ratio. These findings suggest that a reasonable level of sensitivity could be maintained using our partitioning and subsequent aggregation methods.

※ This chapter will be submitted to a peer-reviewed journal for publication.

Chapter 3. Development and Validation of a Predictive Model for Clinical Critical Events in Preterm Infants Admitted to the NICU

3.1. Introduction

In this chapter, we utilized the frameworks presented in preceding chapters to identify predictors for the early detection of major complications, LONS and mortality, in NICU preterm infants. Furthermore, we validated the identified predictors in an external dataset to demonstrate the hypothesis that our proposed frameworks enable the selection of robust relevant features.

In addition, we developed and validated a predictive model to evaluate whether the features identified using the methodologies proposed in the previous chapter contain sufficient information for forecasting clinical events in preterm infants admitted to the NICU. Furthermore, we assessed the applicability of these features in practical predictive modeling to determine their effectiveness in supporting early risk detection and clinical decision-making.

3.2. Methods

3.2.1. Study Design

This study was approved by the Institutional Review Board of Seoul National University Bundang Hospital (IRB No. B-1806-472-106). In this retrospective study,

we used continuous vital sign data recorded from bedside patient monitors in the NICU, as well as demographic information extracted from EMRs. The study population included inborn infants admitted to the NICU at Seoul National University Bundang Hospital (SNUBH) between March 2018 and December 2022, and neonates admitted to the NICU at the University of Virginia (UVA) hospital between January 2009 and December 2019 [35]. To develop a mortality prediction model for low-birth-weight preterm infants, only those for whom continuous monitoring data were available were included in the analysis. Furthermore, to reduce methodological differences from previous studies, the analysis was restricted to infants for whom heart rate and oxygen saturation data were available, consistent with the variables used in prior work [35].

We included preterm infants with GA<32 weeks or birth weight<1,500 grams for this study. Infants were excluded if GA or birth weight data were missing, or if heart rate or oxygen saturation measurements were not available for a minimum of 24 hours.

3.2.2. Data sources

Demographic and clinical data for infants in the NICU at SNUBH were obtained from the hospital's EMR system. Continuous vital sign data were collected using Philips patient monitors. The heart rate data used in this study was obtained from two separate sources, with ECG-derived heart rate and pulse rate measured through pulse oximeter both included in the analysis. Oxygen saturation was measured using pulse oximeter, and respiratory rate was measured through chest impedance monitoring. Blood pressure measurement data consisted of both invasive and non-invasive methods. ECG-derived heart rate, pulse, oxygen saturation, respiratory rate, and invasive blood pressure were stored at 30-second intervals, whereas non-

invasive blood pressure was typically recorded at 30-minute intervals.

In the NICU at the UVA Hospital, ECG-derived heart rate measured by the BedMaster Ex bedside monitoring system was sampled and stored at 0.5 Hz. Similarly, oxygen saturation data measured by the Masimo SET pulse oximetry device were also sampled at 0.5 Hz and used in the analysis [35].

3.2.3. Eligibility criteria and outcome

In this study, the primary outcome was all-cause mortality occurring after the first 24 hours of life. The index time was defined as the designated clinical assessment point. Infants who died during NICU admission were classified into the expired group, with the index time set between 24 and 48 hours prior to death. Infants meeting inclusion criteria who survived to discharge were assigned to the survival group, with their respective index times categorized accordingly. For the identification of predictors and subsequent model development, the prediction execution time was defined as the point at which the predictive model was applied. In cases of sepsis, the index time was established as the earliest date between the blood collection time for a positive blood culture and the initiation of antibiotic therapy within five days of that collection. For these sepsis predictions, the positive class encompassed data from 24 to 48 hours preceding this index time. For all-cause mortality, the index time corresponded to the time of death, with the positive class being defined by data from 24 to 48 hours prior to this established index time.

3.2.4. Predictors

Features for the all-cause mortality predictive model were extracted using the framework previously proposed. We utilized continuous vital signs, specifically heart rate (derived from ECG and pulse oximeter), invasive or non-invasive blood

pressure, and respiratory rate, as inputs for the predictive model. To assess both short-term and long-term effects, data were analyzed across various window sizes: 1, 2, 3, 6, 12, and 24 hours. For feature extraction, continuous vital signs features were generated using all possible combinations of measured vital signs, the various domain-specific feature calculation methods described in Chapter 2, and defined observation window sizes (1, 2, 3, 6, 12, 24 hours). This means that each unique vital sign-feature calculation method pair was applied across every specified observation window size.

3.2.5. Statistical Analysis

Descriptive statistics were employed to summarize baseline characteristics. Normality of continuous data distributions was evaluated using the Kolmogorov–Smirnov test. Variables with normal distribution are presented as means with standard deviations (SD) and compared using two-sided Student’s t tests. For variables not conforming to a normal distribution, medians with interquartile ranges (IQR) were reported and comparisons conducted using the Mann–Whitney U test. Categorical variables were analyzed using the Chi-Squared test or Fisher’s exact test, as appropriate. Variables with more than 50% missing observations were excluded from further analysis. All statistical tests were two-sided, and a p-value of less than 0.05 was considered statistically significant.

3.2.6. Predictive Model Development and Evaluation

We utilized the PyCaret package [124] for the training, optimization, and validation of our predictive models. The models in this study included logistic regression, decision tree classifier, random forest [125], multilayer perceptron (MLP), gradient boosting machine [126], AdaBoost [127], Naïve Bayesian, and Ridge classifier.

Hyperparameter tuning and scaler fitting were performed on the development cohort, with stratified 5-fold cross-validation applied for hyperparameter tuning within this cohort. To assess model robustness, the trained models were evaluated on independent internal and external validation cohorts without further calibration. Model performance was compared using accuracy, area under the receiver operating characteristics curve (AUROC), average precision (AP), recall, precision, and F1 score.

3.3. Results

3.3.1. Study Population

From 436 infants admitted to the SNUBH NICU between March 2018 and December 2022, preterm infants born from March 2018 to June 2021 were allocated to the development cohort, while those born from July 2021 to November 2022 constituted the internal validation cohort. (Figure 3-1 A). For external validation, 1,689 infants met the inclusion criteria were identified from a cohort of 6,837 infants admitted to the UVA NICU between January 2009 and December 2019 (Figure 3-1B).

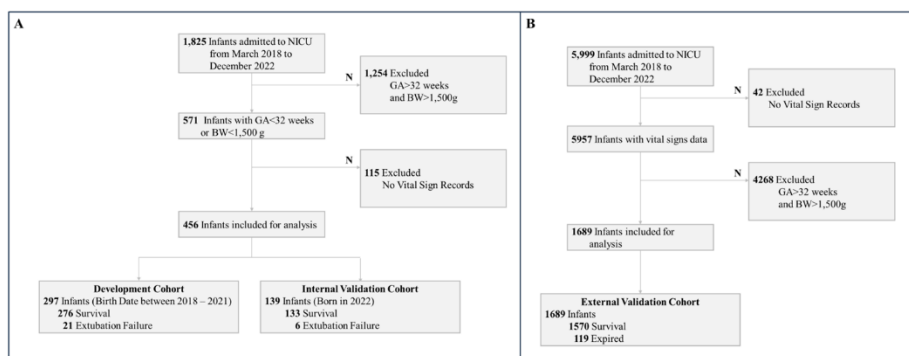


Figure 3-1. Development and internal and external validation cohort to develop and evaluate all-cause predictive models. (A) Development cohort and internal validation cohort. (B) External validation cohort.

Table 3-1 shows the demographic and clinical characteristics of the three study cohorts. The development cohort included 297 patients, of whom 276 survived and 21 died during the admission. Comparing the expired and surviving preterm infants, the expired group demonstrated significantly lower mean GA (25.4 [2.3] weeks) and birth weight (688.7 [273.1] g) compared to the survival group (31.2 [1.9] weeks and 1431.1 [341.5] g, respectively).

The internal validation cohort showed similar demographic trends to the development cohort. Of these 139 preterm infants, only 6 preterm infants expired during the admission. Consistent with findings from the development cohort, the expired group in the internal validation cohort exhibited significantly lower GA (24.3 [0.5] weeks) and birth weight (618.3 [149.6] g) compared to the survival group (GA, 30.9 [2.1] weeks; birth weight, 24.3 [0.5] g). The external validation cohort included 1,563 preterm infants, with 113 infants in the expired group, showing demographic trends similar to the SNUBH preterm infants. In the external validation cohort, the expired group exhibited significantly lower mean [SD] GA (27.1 [3.5] weeks) and birth weight (954.4 [441.2] g) compared to the survival group. Notably, comparing

the expired preterm infants from the UVA with those from SNUBH, the UVA group had slightly higher birth weight and GA.

Table 3-1. Baseline characteristics.

Development Cohort				
Characteristics	All	Survival Group	Expired Group	p-value
Number of infants	297	276	21	
Gestational Age, mean (SD)	30.8 (2.5)	31.2 (1.9)	25.4 (2.3)	<0.001
Birth weight, mean (SD), g	1378.6 (387.0)	1431.1 (341.5)	688.7 (273.1)	<0.001
Gender, n (%)				
Female	141 (47.5)	130 (47.1)	11 (52.4)	0.810
Male	156 (52.5)	146 (52.9)	10 (47.6)	
APGAR 1 min, mean (SD)	5.7 (1.8)	5.7 (1.7)	3.0 (1.5)	0.007
APGAR 5 min, mean (SD)	7.9 (1.3)	7.9 (1.2)	6.2 (1.5)	0.032
APGAR 10 min, mean (SD)	7.7 (0.9)	7.9 (0.7)	6.0 (0.1)	0.036
Internal Validation Cohort				
Characteristics	All	Survival Group	Expired Group	p-value
Number of infants	139	133	6	
Gestational Age, mean (SD)	30.6 (2.5)	30.9 (2.1)	24.3 (0.5)	<0.001
Birth weight, mean (SD), g	1339.4 (429.8)	1372.0 (409.3)	618.3 (149.6)	<0.001
Gender, n (%)				
Female	68 (48.9)	66 (49.6)	2 (33.3)	0.681
Male	71 (51.1)	67 (50.4)	4 (66.7)	
APGAR 1 min, mean (SD)	5.2 (1.8)	5.2 (1.8)	3.0 (1.7)	0.152
APGAR 5 min, mean (SD)	7.6 (1.4)	7.7 (1.3)	4.3 (2.9)	0.177
APGAR 10 min, mean (SD)	6.4 (2.4)	7.0 (1.4)	5.3 (3.8)	0.530
External Validation Cohort				
Characteristics	All	Survival Group	Expired Group	p-value
Number of infants	1689	1570	119	
Gestational Age, mean (SD)	29.1 (3.1)	29.2 (3.0)	27.1 (3.5)	<0.001
Birth weight, mean (SD), g	1259.3 (472.1)	1282.3 (466.5)	954.4 (441.2)	<0.001
Gender, n (%)				
Female	789 (46.7)	741 (47.2)	48 (40.3)	0.177
Male	900 (53.3)	829 (52.8)	71 (59.7)	
APGAR 1 min, mean (SD)	5.2 (2.6)	5.4 (2.6)	3.2 (2.3)	<0.001
APGAR 5 min, mean (SD)	6.9 (2.0)	7.1 (1.9)	5.3 (2.4)	<0.001
APGAR 10 min, mean (SD)	6.7 (1.8)	6.8 (1.7)	5.7 (2.0)	<0.001

3.3.2. Identified Features of Clinical Critical Events

The primary features selected using the proposed framework are detailed in Table 3-2. For heart rate, sample entropy, multiscale sample entropy, permutation entropy, approximate entropy, and absolute sum of change were identified as important features from both ECG and pulse oximeter data. Notably, the expired group consistently exhibited significantly lower mean [SD] values across several entropy measures: approximate entropy (1.112 [0.118] vs. 1.469 [0.139]) and multiscale sample entropy (0.663 [0.355] vs. 1.176 [0.315]). In addition to information-

theoretic methods, decorrelation time selected as a key feature derived from heart rate. Contrary to the entropy-based features, the expired group demonstrated a significantly higher mean [SD] decorrelation time (21.561 [6.236]) compared to the survival group (14.281 [4.440]). This similar tendency was observed in heart rate measured from pulse oximeter. For blood pressure, in contrast to high-frequency vital signs, mid-range, median, and mode were identified as primary indicators, with the expired group consistently exhibiting lower values than the survival group for these features.

Table 3-2. Selected features for all-cause mortality from the development cohort.

Features	Survival	Expired	p-value
HR, absolute sum of changes; 24h	8071 (2787.638)	4161.564 (3436.378)	<0.001
HR, approximate entropy (m=2, r=0.1), 24h	1.469 (0.139)	1.112 (0.118)	<0.001
HR, approximate entropy (m=2, r=0.9), 6h	0.282 (0.094)	0.158 (0.118)	<0.001
HR, decorrelation time, 24h	14.281 (4.440)	21.561 (6.236)	<0.001
HR, multiscale sample entropy (m=2), 24h	1.176 (0.315)	0.663 (0.355)	<0.001
NBP-D, mid-range, 12h	40.280 (6.922)	31.022 (7.671)	<0.001
NBP-D, mode, 24h	38.008 (6.718)	26.116 (7.751)	<0.001
NBP-S, median, 24h	64.954 (6.426)	50.938 (10.077)	<0.001
NBP-S, mode, 6h	60.298 (8.013)	50.325 (14.511)	0.014
Pulse, approximate entropy (m=2, r=0.3), 3h	0.884 (0.194)	0.522 (0.366)	<0.001
Pulse, approximate entropy (m=2, r=0.9), 12h	0.265 (0.079)	0.134 (0.103)	<0.001
Pulse, permutation entropy (d=3, tau=1), 12h	1.722 (0.034)	1.541 (0.163)	<0.001
Pulse, sample entropy, 12h	1.060 (0.267)	0.536 (0.345)	<0.001

Abbreviations: HR, heart rate measured from an electrocardiogram; NBP-D, non-invasive diastolic blood pressure; NBP-S non-invasive systolic blood pressure; Pulse, heart rate measured from a pulse oximeter.

To assess whether the proposed framework selects a consistent feature set during the external validation, we performed the same feature extraction and selection procedure (Table 3-3). The external validation cohort, being larger and containing more data than the SNUBH development cohort, resulted in a greater number of selected features. Consistent with the development cohort, entropy-based features were predominantly extracted from heart rate measurements. Similarly, both entropy and decorrelation time features from SpO₂ were selected, consistent with

findings for sepsis. However, features derived from non-invasive blood pressure, which had a much wider measurement interval, were not identified as key mortality-associated features in the external validation cohort, unlike in the development cohort.

Table 3-3. Selected features for all-cause mortality from the external validation cohort.

Features	Survival	Expired	p-value
HR, approximate entropy (m=2, r=0.9), 3h	0.445 (0.147)	0.230 (0.176)	<0.001
HR, decorrelation time, 6h	11.571 (4.927)	18.876 (5.968)	<0.001
HR, multiscale sample entropy (m=2), 2h	1.597 (0.454)	0.832 (0.654)	<0.001
HR, permutation entropy (d=6, tau=1), 24h	6.113 (0.245)	5.527 (0.694)	<0.001
Pulse, approximate entropy (m=2, r=0.3), 2h	0.932 (0.160)	0.632 (0.263)	<0.001
Pulse, approximate entropy (m=2, r=0.1), 24h	1.548 (0.210)	1.187 (0.346)	<0.001
Pulse, approximate entropy (m=2, r=0.9), 1h	0.498 (0.145)	0.291 (0.206)	<0.001
Pulse, approximate entropy (m=2, r=0.9), 12h	0.419 (0.115)	0.228 (0.144)	<0.001
Pulse, multiscale sample entropy (m=2), 24h	1.387 (0.306)	0.816 (0.446)	<0.001
Pulse, permutation entropy (d=7, tau=1), 6h	6.301 (0.370)	5.986 (0.501)	<0.001
Pulse, sample entropy, 12h	1.385 (0.364)	0.802 (0.586)	<0.001
SpO ₂ , approximate entropy (m=2, r=0.5), 6h	0.667 (0.253)	0.370 (0.240)	<0.001
SpO ₂ , approximate entropy (m=2, r=0.5), 12h	0.665 (0.238)	0.360 (0.223)	<0.001
SpO ₂ , approximate entropy (m=2, r=0.7), 6h	0.474 (0.182)	0.281 (0.185)	<0.001
SpO ₂ , approximate entropy (m=2, r=0.7), 12h	0.457 (0.185)	0.244 (0.149)	<0.001
SpO ₂ , approximate entropy (m=2, r=0.7), 24h	0.440 (0.166)	0.242 (0.136)	<0.001
SpO ₂ , decorrelation time, 24h	7.519 (6.099)	16.145 (8.420)	<0.001
SpO ₂ , Fourier entropy (bins=5), 6h	0.419 (0.216)	0.274 (0.246)	0.001

Abbreviations: HR, heart rate measured from an electrocardiogram; Pulse, heart rate measured from a pulse oximeter; SpO₂, oxygen saturation.

For LONS, features extracted from SpO₂ were predominantly selected as key indicators (Table 3-4). Consistent with findings for all-cause mortality, the heart rate decorrelation time was significantly longer in the sepsis group (2.448 [1.995]) compared to the control group. Entropy-based features were identified exclusively from pulse oximetry device, showing lower entropy in the sepsis group relative to controls.

Table 3-4. Selected features for late onset sepsis from the development cohort.

Features	Control	Sepsis	p-value
HR, decorrelation time; 1h	1.3557 (1.565)	2.448 (1.995)	<0.001
Pulse, approximate entropy (m=2, r=0.5), 6h	0.643 (0.189)	0.454 (0.213)	<0.001
Pulse, decorrelation time, 1h	4.731 (3.889)	8.300 (5.827)	<0.001
Pulse, sample entropy, 12h	1.091 (0.286)	0.850 (0.362)	<0.001
SpO ₂ , autocorrelation (mean), 2h	0.052 (0.141)	0.131 (0.160)	0.002
SpO ₂ , b-index, 12h	0.029 (0.038)	0.057 (0.048)	<0.001
SpO ₂ , binned entropy, 1h	0.947 (0.540)	1.340 (0.512)	<0.001
SpO ₂ , decorrelation time, 2h	0.865 (0.753)	1.589 (0.767)	<0.001
SpO ₂ , decorrelation time, 3h	0.859 (1.353)	2.218 (2.436)	<0.001
SpO ₂ , decorrelation time, 6h	0.712 (0.618)	2.638 (3.407)	<0.001
SpO ₂ , decorrelation time, 12h	3.082 (3.189)	7.394 (5.882)	<0.001
SpO ₂ , Fourier entropy (bins=100), 2h	2.656 (0.831)	1.970 (0.785)	<0.001
SpO ₂ , Fourier entropy (bins=100), 6h	2.941 (0.521)	2.117 (0.805)	<0.001
SpO ₂ , Gibb's index (m=2), 6h	0.657 (0.227)	0.816 (0.183)	<0.001
SpO ₂ , index mass quantile (Q=0.6), 24h	0.601 (0.092)	0.599 (0.004)	<0.001
SpO ₂ , ranvr, 3h	0.008 (0.006)	0.014 (0.008)	<0.001
SpO ₂ , cross spectral density (Welch's method), 12h	71.981 (85.121)	270.278 (253.700)	<0.001
SpO ₂ , vmr, 3h	0.042 (0.043)	0.168 (0.154)	<0.001
SpO ₂ , winsorized mean, 12h	98.664 (1.109)	96.804 (1.925)	<0.001

Abbreviations: HR, heart rate measured from an electrocardiogram; Pulse, heart rate measured from a pulse oximeter; SpO₂, oxygen saturation.

From SpO₂ data, features related to qualitative variation and those derived from decorrelation time methods were chosen. While the control group exhibited remarkably uniform decorrelation times across observation windows of 1, 2, 3, and 6 hours prior to the evaluation time (index time, t=0), the sepsis group demonstrated a progressive increase in decorrelation time as the observation window expanded. Furthermore, the Gibb's index was significantly higher in the sepsis group.

To evaluate the correlation among the 247,000 continuous vital sign features generated in this study, specifically among the selected features, we calculated an absolute correlation map (Figure 3-2, Figure 3-3, Figure 3-4, Figure 3-5). In the all-cause mortality correlation matrix, we observed that pulse oximeter-derived heart rate, which often excluded from analysis due to its perceived lower reliability, exhibited a correlation with all-cause mortality similar to that of ECG-based heart rate. A comparable correlation pattern was also evident for these features in relation to late-onset sepsis.

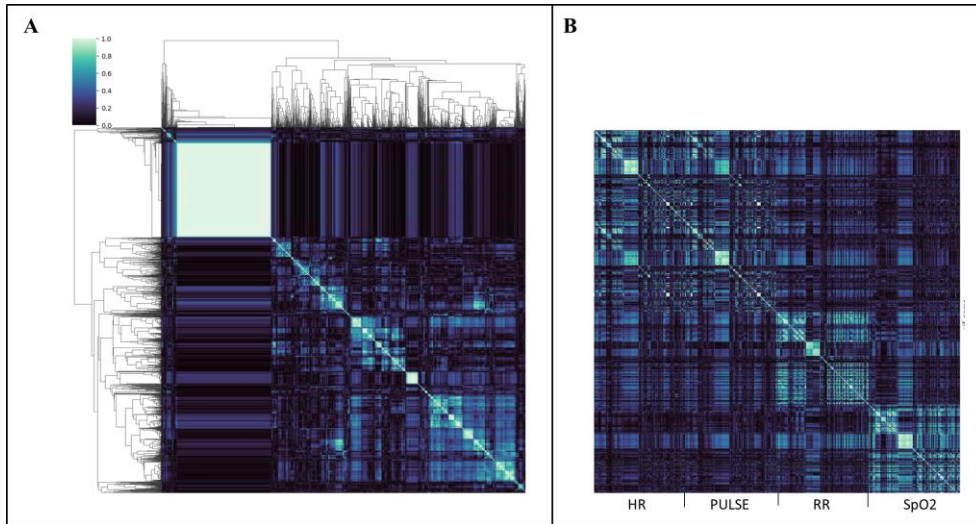


Figure 3-2. Absolute correlation matrix diagram in all-cause mortality (A) Features clustering dendrogram, with leaves representing individual features and nodes indicating clusters. (B) Features grouped by vital sign type.

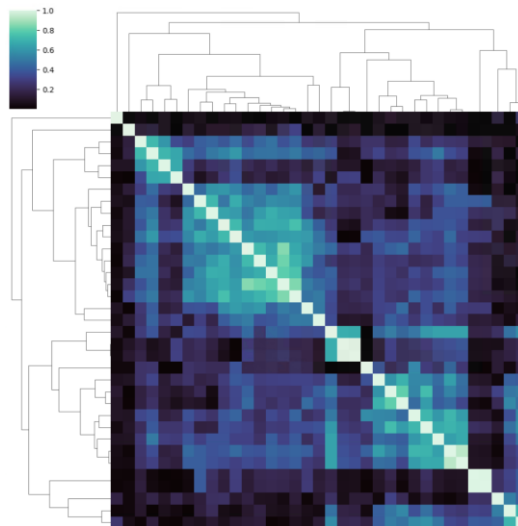


Figure 3-3. Absolute correlation matrix diagram of selected feature in all-cause mortality.

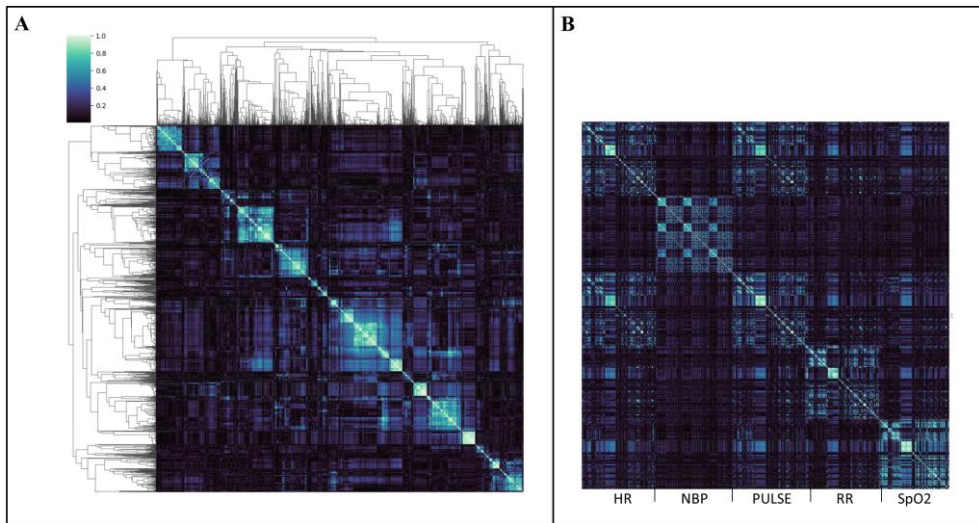


Figure 3-4. Absolute correlation matrix in late onset sepsis. (A) Features clustering dendrogram, with leaves representing individual features and nodes indicating clusters. (B) Features grouped by vital sign type.

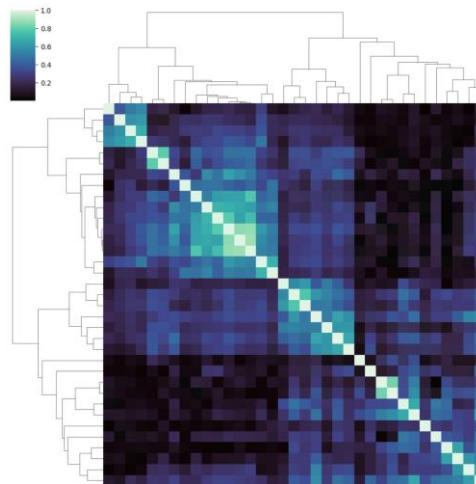


Figure 3-5. Absolute correlation diagram of selected features in late onset sepsis.

3.3.1. Predictive Models Performance

To assess model performance and robustness, we conducted both the internal validation cohort and the external validation cohort using the UVA dataset. Our models demonstrated high performance across both validation cohorts (Table 3-5, Table 3-6, Table 3-7, Table 3-8).

In the internal validation cohort, most models achieved a mean AUROC of over 0.800, indicating excellent performance. Similarly, high performance was observed during the external validation. The Extra Trees Classifier showed the top performance, exhibiting a higher mean AUROC (0.865; 95% CI, 0.864 – 0.866) compared to other models. Notably, despite differences in demographics, data quality, and sampling rates between the two datasets, we observed the generalizability of the models.

Table 3-5. Performance of predictive models using the proposed continuous vital sign analysis method frameworks for clinical critical events.

	Internal Validation Cohort					
Classifier Metrics	Accuracy	AUROC (CI 95%)	AP	Recall	Precision	F1 Score
Proposed Analytic Methods						
Logistic Regression	0.929	0.842 (0.939-0.845)	0.441	0.400	0.423	0.412
Naïve Bayes	0.870	0.818 (0.815-0.822)	0.347	0.636	0.267	0.376
Random Forest Classifier	0.919	0.856 (0.853-0.859)	0.506	0.522	0.386	0.444
Extra Trees Classifier	0.926	0.879 (0.877-0.859)	0.506	0.504	0.418	0.457
Ada Boost Classifier	0.933	0.861 (0.858-0.865)	0.505	0.421	0.457	0.439
Gradient Boosting Classifier	0.920	0.851 (0.847-0.854)	0.448	0.470	0.381	0.421
Decision Tree Classifier	0.862	0.618 (0.619-0.625)	0.140	0.339	0.177	0.233
MLP Classifier	0.892	0.745 (0.741-0.749)	0.274	0.405	0.262	0.318

Abbreviations: AP, average precision; AUROC, area under the receiver operating characteristic curve; MLP, multi-layer perceptron.

Table 3-6. Performance of predictive models using the proposed continuous vital sign analysis method frameworks for all-cause mortality.

	External Validation Cohort					
Classifier Metrics	Accuracy	AUROC (CI 95%)	AP	Recall	Precision	F1 Score
Proposed Analytic Methods						
Logistic Regression	0.759	0.792 (0.788-0.797)	0.294	0.665	0.158	0.255
Naïve Bayes	0.068	0.501 (0.500-0.502)	0.084	0.994	0.062	0.117
Random Forest Classifier	0.907	0.837 (0.833-0.841)	0.419	0.411	0.311	0.354
Extra Trees Classifier	0.943	0.873 (0.869-0.877)	0.539	0.382	0.564	0.379
Ada Boost Classifier	0.832	0.827 (0.823-0.830)	0.343	0.559	0.198	0.292
Gradient Boosting Classifier	0.920	0.824 (0.819 -0.829)	0.403	0.391	0.367	0.379
Decision Tree Classifier	0.868	0.632 (0.627-0.637)	0.145	0.368	0.198	0.257
MLP Classifier	0.516	0.677 (0.672-0.682)	0.145	0.753	0.091	0.162
Abbreviations: AP, average precision; AUROC, area under the receiver operating characteristic curve; MLP, multi-layer perceptron.						

Abbreviations: AP, average precision; AUROC, area under the receiver operating characteristic curve; MLP, multi-layer perceptron.

Table 3-7. Performance of real-time predictive models using the proposed continuous vital sign analysis method frameworks for clinical critical events.

	Internal Validation Cohort					
Classifier Metrics	Accuracy	AUROC (CI 95%)	AP	Recall	Precision	F1 Score
Proposed Analytic Methods						
Logistic Regression	0.942	0.852 (0.852-0.853)	0.416	0.400	0.423	0.412
Naïve Bayes	0.876	0.820 (0.819-0.821)	0.271	0.636	0.267	0.376
Random Forest Classifier	0.944	0.872 (0.871-0.873)	0.441	0.522	0.386	0.444
Extra Trees Classifier	0.946	0.891 (0.891-0.892)	0.468	0.504	0.418	0.457
Ada Boost Classifier	0.933	0.865 (0.865-0.866)	0.445	0.421	0.457	0.439
Gradient Boosting Classifier	0.933	0.862 (0.862-0.863)	0.402	0.470	0.381	0.421
Decision Tree Classifier	0.862	0.617 (0.617-0.618)	0.101	0.339	0.177	0.233
MLP Classifier	0.900	0.757 (0.756-0.758)	0.227	0.405	0.262	0.318

Abbreviations: AP, average precision; AUROC, area under the receiver operating characteristic curve; MLP, multi-layer perceptron.

Table 3-8. Performance of real-time predictive models using the proposed continuous vital sign analysis method frameworks for all-cause mortality.

	External Validation Cohort					
Classifier Metrics	Accuracy	AUROC (CI 95%)	AP	Recall	Precision	F1 Score
Proposed Analytic Methods						
Logistic Regression	0.759	0.758 (0.757-0.760)	0.216	0.665	0.158	0.255
Naïve Bayes	0.068	0.500 (0.499-0.500)	0.062	0.994	0.062	0.117
Random Forest Classifier	0.909	0.837 (0.836-0.838)	0.353	0.405	0.317	0.355
Extra Trees Classifier	0.944	0.865 (0.864-0.866)	0.471	0.372	0.577	0.452
Ada Boost Classifier	0.832	0.807 (0.807-0.808)	0.264	0.559	0.198	0.221
Gradient Boosting Classifier	0.920	0.813 (0.812 -0.814)	0.350	0.391	0.367	0.379
Decision Tree Classifier	0.868	0.634 (0.633-0.635)	0.112	0.368	0.198	0.257
MLP Classifier	0.516	0.663 (0.663-0.664)	0.104	0.753	0.091	0.162

Abbreviations: AP, average precision; AUROC, area under the receiver operating characteristic curve; MLP, multi-layer perceptron.

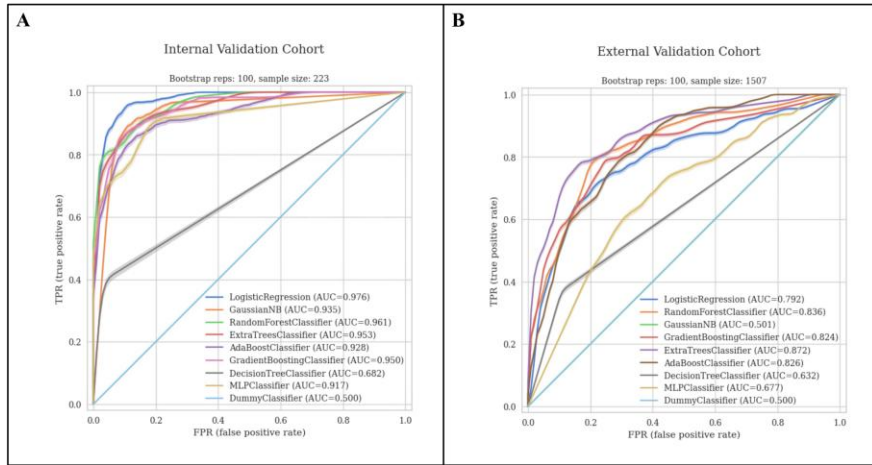


Figure 3-6. Clinical critical event predictive models AUROC performance: (A) Internal validation cohort, (B) External validation cohort.

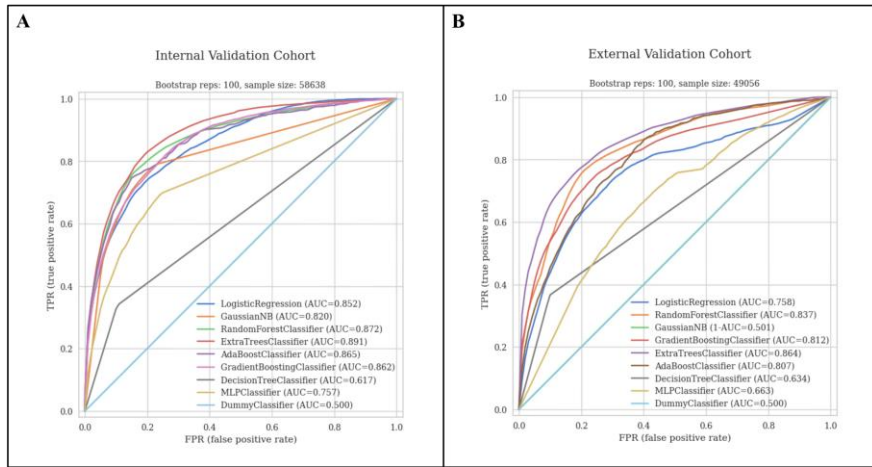


Figure 3-7. Real-Time clinical critical event predictive models AUROC performance (A) Internal validation cohort, (B) External validation cohort.

3.4. Discussion

This study identified features associated with all-cause mortality and LONS in the admitted preterm infants using continuous vital sign data, subsequently developing

the predictive model with high performance and demonstrated generalizability in an external validation cohort. The Extra Trees model demonstrated the top discrimination performance, achieving an AUROC of 0.891 (95% CI, 0.891–0.892) in the internal validation cohort and 0.865 (95% CI, 0.864–0.866) in the external validation cohort, evaluating robust performance across both cohorts. Furthermore, the selected features were consistent with existing indicators and revealed novel utility from other domains.

Many studies have focused on identifying key risk factors and predictors for mortality (including sudden infant deaths) and clinical deterioration in vulnerable preterm infants within the NICU. This often involves integrating these insights with artificial intelligence technologies for early detection. However, despite achieving high performance in internal validation, most studies do not sufficiently address external validation [64, 128]. Consequently, fundamental questions regarding reproducibility and generalizability remain unaddressed, significantly impeding the implementation of these predictive models in routine NICU clinical practice. In this study, we demonstrated that our predictive model achieved high performance and generalizability when applied to continuous vital sign data from an external institution, utilizing the same feature calculation methods as the development cohort without additional processing or calibration. This outcome suggests that continuous vital sign-based features might mitigate limitations inherent to EMR-based predictive models, which often suffer from significant inter-institutional and inter-clinician variability [40]. Furthermore, the consistent high performance across various patient monitor manufacturers suggests the broad adaptability of these models for predicting preterm infant deterioration without requiring additional equipment.

This study identified low heart rate entropy and decorrelation time as key

features. While entropy-based heart rate features were frequently cited in research, their utility has often been limited by significant inter-institutional variability, making them robust only within single institutions [27, 40]. We validated the importance of low heart rate sample entropy, a consistently identified key feature in prior work, while simultaneously introducing heart rate decorrelation time, which demonstrated potential for more robust applications. Notably, despite the known heterogeneity in NICU settings leading to substantial inter-institutional variability in sample entropy values, both institutions in this study showed consistently high contributions from this feature. Such discrepancies have often been attributed to demographic differences and measurement equipment. However, most previous research on entropy-based features has reduced data resolution through techniques like segmenting vital signs, random sampling, or using grand means/medians [30, 34, 35, 39]. This was primarily due to the computational demands of large-volume continuous vital signs and the high resource requirements for approximate and sample entropy. These approaches, while computationally efficient, limit reproducibility due to disparate aggregation methods and risk missing subtle preterm symptoms by downsampling high-frequency data. Our study mitigated these issues by directly utilizing up to 24 hours of continuous vital signs, identifying sample entropy as a highly interpretable and discriminative indicator when derived from high-resolution, high-frequency data collected via bedside patient monitors. This was further supported by consistent findings across both pulse oximetry and ECG measurements.

✂ This chapter will be submitted to a peer-reviewed journal for publication.

Chapter 4. Predictive Modeling of Extubation Readiness in Preterm Infants Using Real-Time Physiological Data

4.1. Introduction

Preterm infants frequently require endotracheal intubation and invasive mechanical ventilation during the early postnatal period, primarily due to pulmonary immaturity, insufficient central respiratory drive, and surfactant deficiency. Although mechanical ventilation is an important component of neonatal intensive care, prolonged use is associated with increased risks of bronchopulmonary dysplasia, neurodevelopmental sequelae, and all-cause neonatal mortality [129, 130].

In this chapter, we developed and validated a predictive model for extubation success within 24 hours in preterm infants, using the proposed continuous vital sign feature analysis framework. This study aimed to validate the hypothesis that features derived from continuous vital sign time series data provide additional, clinically meaningful information not detected by conventional traditional EMR-based features. To evaluate this, model performance was compared against the previously developed EMR-based NeXT Predictor [131] under identical experimental conditions, to assess the potential superiority of the proposed approach.

4.2. Methods

4.2.1. Study Design

This study was approved by the Institutional Review Board of SNUBH (IRB No. X-2205-759-901). In this retrospective study, we used continuous vital sign data recorded from bedside patient monitors in the NICU, as well as demographic information extracted from EMR. The study population included inborn infants admitted to the NICU at SNUBH between March 2018 and December 2022.

4.2.2. Eligibility Criteria and Outcome

This study enrolled preterm infants born at less than 32 weeks of GA who were managed with mechanical ventilation via an endotracheal tube and underwent their planned extubation attempt prior to 36 weeks of postmenstrual age (PMA). Infants with major congenital anomalies or structural airway abnormalities were excluded, as were those extubated within 6 hours of initial intubation. Only infants intubated for more than 6 hours were included, thereby excluding cases of procedural intubation. Unplanned extubations were excluded. Planned extubation was set as the index ($t=0$) for outcome evaluation. The observation period was from NICU admission to index data collection.

The primary outcome was the success or failure of the planned extubation in preterm infants. Extubation success was defined as the absence of reintubation within 72 hours following the planned extubation. Reintubations occurring within 10 minutes of extubation were excluded from the analysis, as it was not feasible to reliably discriminate between true extubation failure and events such as unplanned self-extubation or a misplaced endotracheal tube.

4.2.3. Predictors

In this study, predictors for modeling extubation readiness in preterm infants were derived using the continuous vital sign analysis framework described in the preceding chapter. The dataset included all routinely collected vital signs from patient monitors, including heart rate, pulse, oxygen saturation, blood pressure, and respiratory rate. To assess both short- and long-term physiological dynamics, feature observation windows were defined at intervals of 1, 2, 3, 6, 12, and 24 hours. For each window, all possible combinations of time intervals, feature extraction methods, and vital sign modalities were generated to construct a comprehensive feature set. In cases where more than 50% of data within a given observation window were missing, the corresponding feature was treated as missing value.

In previous NeXT-Predictor study, clinical and physiological data were retrospectively analyzed to identify potential predictors of extubation failure [131]. Data was collected prior to extubation and, when applicable, prior to reintubation. Routinely recorded vital signs—including heart rate, respiratory rate, body temperature, oxygen saturation, and blood pressure—were included. Candidate predictors derived from GA, birth weight, PMA at the time of extubation, male sex, pre-extubation arterial blood gas measurements (pH and partial pressure of carbon dioxide [$p\text{CO}_2$]), and ventilator settings such as the fraction of inspired oxygen (FiO_2), positive end-expiratory pressure (PEEP), mean airway pressure (MAP), and ventilator respiratory rate setting. In addition, respiratory indices, including the $\text{SpO}_2/\text{FiO}_2$ (SF) ratio [132], Respiratory Rate Oxygenation (ROX) index [133], respiratory severity score (RSS) were evaluated. Predictive features were derived by applying time-domain methods to vital sign variables that were periodically measured from admission to the index time point (Table 4-1).

Missing data from ventilator setting was imputed using the last observation carried forward (LOCF) method. This approach was selected for two reasons. First, it helps preserve the original distribution of the data, which was critical given that this study focused on the variability of physiological parameters; minimizing the introduction of artificial bias or distortion of statistical properties was essential. Second, most missing values were for ventilator settings, which EMR recorded only when clinicians made substantial adjustments. In contrast, vital signs were continuously and automatically recorded, resulting in minimal data loss. Given this clinical context, we considered it reasonable to assume that unrecorded ventilator values likely remained consistent with prior entries, and any potential changes during the gap were presumed to be clinically insignificant.

Table 4-1. Feature extraction method used in NExT predictor model [131].

Time domain methods	Formula
Mean	$\bar{x} = \frac{\sum_{i=1}^N X_i}{N}$
Standard Deviation	$s = \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2 / N}$
Outlying	$\frac{\sum x_{outlier}}{\sum x_i}$ where $x_{outlier} \notin A, A = \{x x_{0.25} - 1.5 \times IQR < x < x_{0.85} + 1.5 \times IQR\}$
Sequential difference Mean	$\bar{x} = \frac{\sum_{i=1}^N \Delta x_i}{N}$ Lag difference (lag-1 difference)
Sequential difference of Standard Deviation	$s = \sqrt{\sum_{i=1}^N (\Delta Y_i - \overline{\Delta Y})^2 / N}$
Sequential difference Outlying	$\frac{\sum \Delta x_{outlier}}{\sum \Delta x_i}$ where $\Delta x_{outlier} \notin A, A = \{\Delta x \Delta x_{0.25} - 1.5 \times IQR < \Delta x < \Delta x_{0.85} + 1.5 \times IQR\}$
Trend	nonparametric Mann-Kendall test
Periodicity	Fourier coefficient
Randomness	$r_k = \frac{\sum_{i=1}^{N-k} (Y_i - \bar{Y})(Y_{i+k} - \bar{Y})}{\sum_{i=1}^N (Y_i - \bar{Y})^2}$

4.2.4. Statistical Analysis

Baseline characteristics were analyzed using descriptive statistics. The distribution of continuous variables was assessed for normality using the Kolmogorov–Smirnov test. Variables following a normal distribution were reported as mean with standard deviation (SD) and were compared using two-tailed Student’s *t* tests. Non-normally distributed variables were reported as median with interquartile range (IQR) and were analyzed using the Mann–Whitney *U* test. Categorical variables were compared using either the Chi-Square test or Fisher’s exact test, as appropriate. Variables with more than 50% missing data were excluded from the analysis.

Within the proposed continuous vital sign analysis framework, we conducted 200 iterations of bootstrap resampling with replacement to identify candidate predictive features. To address the high dimensionality of the resulting feature set relative to the sample size and to mitigate multicollinearity, variance inflation factor (VIF) analysis was employed. Features with the highest VIF values were iteratively removed, beginning with the most redundant, until a parsimonious and stable feature set was finalized for model development.

In in NExT-Predictor, Statistical analyses were conducted using the statsmodels [94] and tableone [134] Python libraries. Propensity score matching was applied to identify candidate features that showed statistically significant differences between the outcome and control groups. Univariable analyses were used to estimate adjusted odds ratios (ORs) and marginal effects for each predictor, with GA and birth weight included as covariates due to their known influence on extubation outcomes in preterm infants. Continuous variables were assessed at the time of extubation using univariable methods, such as Student’s *t* test, while categorical variables were evaluated using the Chi-Square test for baseline comparisons.

4.2.5. Predictive Model Development and Evaluation

All classification models were developed using the PyCaret machine learning library (v3.0) [124], which provided a unified API for model training, preprocessing, and evaluation. The models included logistic regression, decision tree (DT), extra tree forest(ET), random forest (RF), gradient boosting machine (GBM), stochastic gradient descent (SGD) classifier, naïve Bayes (NB), and extreme gradient boosting (XGBoost). PyCaret’s classification module was used to automate pipeline construction, including standardization, imputation, and cross-validation. Hyperparameters were optimized using grid search within PyCaret’s built-in tuning function. This approach enabled consistent preprocessing and fair comparison across model types representing linear, probabilistic, and ensemble-based learning algorithms. The development cohort was used for model development, including stratified 10-fold cross-validation and hyperparameter tuning via grid search, with F1 score as the primary optimization metric. The internal validation set, held out from the initial split, was used only for final performance evaluation to ensure an unbiased assessment. Details of the hyperparameter configurations and search spaces are provided in Table 4-2. To avoid data leakage, all model tuning and selection were performed strictly to the training data in the development cohort.

Table 4-2. Hyperparameters for developing prediction models.

Classifier	Parameter name	Parameter range
Logistic Regression	C	0.001 – 1,000
	Penalty	None, L1, L2
	Class Weight	None, Balanced
Random Forest	number of estimators	50, 100, 300, 600, 1000, 2000
	Maximum Depth	3, 5, 10, Inf
	Criterion	gini coefficient, entropy
Gradient Boosting	Loss	Deviance, Exponential
	Learning Rate	0.01 – 1.0
	number of estimators	50, 100, 300, 600, 1000, 2000
	Maximum Depth	3, 5, 10
	Learning Rate	0.01, 0.1, 0.3, 1.0
XGBoost	number of estimators	50, 100, 300, 600, 1000, 2000
	Maximum Depth	3, 5, 10, 15
	Minimum Child Weight	1,3,5
	Loss	Modified Huber, Log
Stochastic Gradient Decent	Alpha	0.0001, 0.00001, 0.000001
	Penalty	L2, Elasticnet
	Criterion	Gini Coefficient, Entropy
Decision Tree	Maximum Depth	2, 4, 6, 8, 10, 12
	Alpha	0.001 – 1,000,000
Complement Naïve Bayesian	Fit Prior	1, 0
	Normalization	1, 0

Model discrimination was evaluated using multiple performance metrics, including accuracy, the area under the receiver operating characteristic curve (AUROC), the area under the precision–recall curve (AUPRC), positive predictive value (PPV), and negative predictive value (NPV).

4.3. Results

4.3.1. Study Population

A total of 253 preterm infants met the inclusion criteria and were enrolled during the study period. Of these, 185 infants (73%), born between March 2018 and December 2021, were assigned to the development cohort, and 68 infants (27%), born in 2022, were allocated to the internal validation cohort. Extubation failure occurred in 69 infants (26%) across the entire study population—53 infants (29%) in the development cohort and 16 infants (24%) in the internal validation cohort.

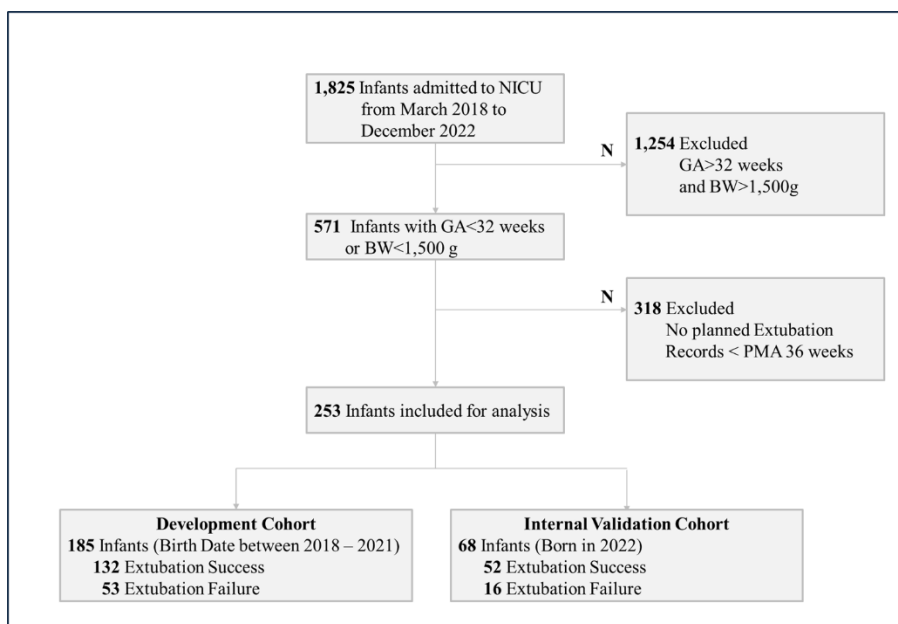


Figure 4-1. Development and internal validation cohort to identify predictors and develop predictive models.

Table 4-3, Table 4-4 show a summary of the baseline characteristics within the development cohort and the internal validation cohort, respectively. In the

development cohort, the mean (SD) GA was 27.3 (2.5) and 29.2 (2.7) weeks in the extubation and the extubation success group, respectively. The mean (SD) birth weight was 896.7g (367.5) and 1,203.6g (472.9), respectively. There were no statistically significant differences in ventilator settings at the time of extubation as determined by the clinicians between the two groups. FiO₂ was 0.30 (0.05) in the extubation success group and 0.32 (0.06) in the failure group; PEEP was 5.2 (1.1) and 5.3 (1.0) cmH₂O; MAP was 9.6 (1.5) vs 9.7 (1.6) cmH₂O, respectively.

In the internal validation cohort, the mean (SD) GA was 25.5 (1.5) and 29.5 (2.7) weeks in the extubation and the extubation success group. The mean (SD) of birth weight was 896.7g (367.5) and 1,203.6g (472.9), consistent with baseline characteristics observed in the development cohort. Similarly, no statistically significant differences in ventilator settings at the time of extubation, as determined by the clinicians, were found between the two groups.

Table 4-3. Baseline characteristics in the development cohort.

Characteristics	Development Cohort			p-value
	All	Extubation Success Group	Extubation Failure Group	
Number of infants	185	132	53	
Gestational Age, mean (SD), weeks	28.7 (2.8)	29.2 (2.7)	27.3 (2.5)	<0.001
Birth weight, mean (SD), g	1135.0 (468.7)	1203.6 (472.9)	896.7 (357.5)	<0.001
Gender, n (%)				
Female	118 (43.2)	92 (43.4)	26 (42.6)	1.000
Male	155 (56.8)	120 (56.6)	35 (57.4)	
PMA at extubation (weeks)	31.3 (2.1)	31.5 (2.0)	30.7 (2.3)	0.004
Ventilation Variables				
FiO ₂ , mean (SD),	0.26 (0.09)	0.24 (0.07)	0.32 (0.11)	0.001
PEEP, mean (SD), cm H ₂ O	5.8 (0.8)	5.6 (0.7)	6.0 (0.8)	0.001
MAP, mean (SD), cm H ₂ O	8.8 (1.6)	8.5 (1.3)	9.4 (2.0)	0.001
Frequency mean (SD), rpm	31.6 (6.4)	31.5 (6.0)	32.1 (7.4)	0.511
Ventilation Variables after Post Extubation				
FiO ₂ , mean (SD),	0.30 (0.09)	0.27 (0.07)	0.35 (0.10)	<0.001
PEEP, mean (SD), cm H ₂ O	5.9 (0.7)	5.7 (0.7)	6.2 (0.7)	<0.001
MAP, mean (SD), cm H ₂ O	9.1 (1.9)	8.5 (1.6)	9.5 (2.0)	0.002
Frequency mean (SD), rpm	34.1 (9.1)	34.1 (9.2)	34.1 (9.0)	0.992

Abbreviations: FiO₂, fraction of inspired oxygen; MAP, mean airway pressure; PEEP, positive end-expiratory pressure PMA, postmenstrual age.

Table 4-4. Baseline characteristics in the internal validation cohort.

Characteristics	Internal Validation Cohort			p-value
	All	Extubation Success Group	Extubation Failure Group	
Number of infants	68	52	16	
Gestational Age, mean (SD), weeks	29.0 (2.7)	29.8 (2.5)	26.4 (1.9)	<0.001
Birth weight, mean (SD), g	1139.3 (459.2)	1240.2 (471.8)	711.2 (187.3)	<0.001
Gender, n (%)				
Female	36 (52.9)	27 (51.9)	9 (56.2)	0.987
Male	32 (47.1)	25 (48.1)	7 (43.8)	
PMA at extubation (weeks)	30.9 (1.9)	31.1 (1.9)	30.5 (2.0)	0.145
Ventilation Variables				
FiO ₂ , mean (SD),	0.28 (0.11)	0.24 (0.10)	0.35 (0.10)	<0.001
PEEP, mean (SD), cm H ₂ O	6.4 (0.9)	6.1 (0.8)	7.0 (0.7)	<0.001
MAP, mean (SD), cm H ₂ O	10.7 (2.4)	10.3 (2.5)	11.1 (2.3)	0.152
Frequency mean (SD), rpm	28.0 (4.6)	28.0 (4.6)	28.1 (4.6)	0.914
Ventilation Variables after Post Extubation				
FiO ₂ , mean (SD),	0.30 (0.09)	0.27 (0.07)	0.37 (0.08)	<0.001
PEEP, mean (SD), cm H ₂ O	6.5 (0.9)	6.2 (0.9)	7.1 (0.7)	<0.001
MAP, mean (SD), cm H ₂ O	11.6 (3.2)	11.8 (4.3)	11.5 (2.5)	0.786
Frequency mean (SD), rpm	31.3 (6.2)	34.6 (5.6)	30.2 (6.0)	0.024

Abbreviations: FiO₂, fraction of inspired oxygen; MAP, mean airway pressure; PEEP, positive end-expiratory pressure PMA, postmenstrual age.

4.3.2. Predictors of Extubation Failure

To analyze the selected continuous vital sign features relevant to extubation readiness, their mean and standard deviation were calculated at a simulated extubation assessment time point. Variables significantly associated with extubation failure are detailed in Table 4-5. At the time of extubation assessment, heart rate decorrelation time was notably longer in the extubation failure group (14.857 [5.639]) compared to the extubation success group (12.846 [4.857]). Heart rate skewness also consistently exhibited negative values across the entire observation window for the extubation failure group. Similarly, pulse decorrelation time, as measured by pulse oximetry, was prolonged in the extubation failure group. Furthermore, oxygen saturation (SpO₂) in the extubation failure group demonstrated greater variation and lower mean values.

Table 4-5. Clinically relevant features identified for extubation.

Features	Extubation Success	Extubation Failure	p-value
HR, decorrelation time; 12h, mean (SD)	12.846 (4.857)	14.857 (5.639)	0.002
HR, quantile (q=0.9), 1h	162.652(14.326)	168.785 (14.,925)	0.001
HR, quantile (q=0.75), 1h	154.961 (13.762)	163.500 (14.654)	<0.001
HR, skew, 6h	0.195 (1.836)	-0.740 (2.267)	<0.001
HR, skew, 1h	0.403 (1.402)	-0.351 (1.438)	<0.001
HR., skew, 24h	-0.049 (1.787)	-0.848 (2.112)	0.001
Pulse, decorrelation time, 24h	24.483 (10.234)	29.276 (10.058)	<0.001
Pulse, decorrelation time, 12h	4.983 (6.506)	7.367 (8.353)	0.009
SpO ₂ , approximate entropy (m=2, r=0.1), 12h	0.858 (0.366)	1.088 (0.288)	<0.001
SpO ₂ , harmonic mean, 3h	97.760 (1.951)	95.914 (2.620)	<0.001
SpO ₂ , mode absolute deviation, 1h	1.636 (1.489)	2.676 (1.631)	<0.001
SpO ₂ , mode absolute deviation, 24h	1.742 (1.144)	2.627 (1.299)	<0.001
SpO ₂ , permutation entropy (d=7, tau=1), 2h	3.486 (1.433)	4.295 (0.988)	<0.001
SpO ₂ , q statistic, 6h	77.012 (56.097)	125.697 (67.160)	<0.001

Abbreviations: HR, heart rate measured from an electrocardiogram; Pulse, heart rate measured from a pulse oximeter; SpO₂, oxygen saturation.

4.3.3. Model Performance

Model training results based on the proposed continuous vital sign analytics framework are summarized in Table 4-6. Among the candidate models, the logistic regression classifier using continuous vital sign-derived features achieved the highest discriminatory performance in the internal validation cohort, with an AUROC of 0.976 (95% CI, 0.974–0.978), indicating high discrimination and outperforming all other models [135] (Table 4-6).

Table 4-6. Performance of predictive models utilizing the proposed continuous vital sign analysis methodology.

Classifier Metrics	Internal Validation Cohort					
	Accuracy	AUROC (CI 95%)	AP	Recall	Precision	F1 Score
Proposed Analytic Methods						
Logistic Regression	0.928	0.976 (0.974-0.978)	0.901	0.872	0.815	0.842
Naïve Bayes	0.861	0.935 (0.932-0.938)	0.729	0.908	0.594	0.719
Random Forest Classifier	0.864	0.961 (0.959-0.963)	0.868	0.833	0.761	0.795
Extra Trees Classifier	0.838	0.953 (0.950-0.955)	0.864	0.816	0.792	0.804
Ada Boost Classifier	0.860	0.928 (0.924-0.932)	0.753	0.778	0.830	0.803
Gradient Boosting Classifier	0.916	0.950 (0.947-0.952)	0.841	0.778	0.830	0.803
Decision Tree Classifier	0.847	0.682 (0.675-0.689)	0.456	0.403	0.805	0.537
MLP Classifier	0.899	0.917 (0.913-0.922)	0.806	0.681	0.821	0.775

Abbreviations: AP, average precision; AUROC, area under the receiver operating characteristic curve; MLP, multi-layer perceptron.

4.4. Discussion

In this study, we developed a predictive model using continuously recorded vital signs obtained directly from patient monitors, combined with the proposed analytic framework for feature extraction and selection. The model demonstrated superior performance compared with conventional EMR-based models. Evaluation metrics, including the AUROC, F1 score, and accuracy, indicated consistently high discriminative ability and calibration across both the development and internal validation cohorts.

The extubation success rate from our dataset's source institution, using the more recent internal validation cohort, showed the NExT-Predictor model achieving an AUROC of 0.752. This performance surpasses that of Gupta et al.[130]'s predictive model but remains relatively lower compared to models utilizing continuous vital sign-derived predictors. In contrast, models based on continuous vital sign data demonstrated exceptionally high AUROC, average precision, and strong calibration. We hypothesized this discrepancy occurs because models heavily reliant on EMR data are more vulnerable to clinician input variability and human error. The demand for continuous physiological predictors grows because EMR data input is clinician-dependent, with documentation frequency increasing in deteriorating or vulnerable infants. This introduces potential data bias. Therefore, we expect that continuous vital sign-based predictors could effectively address the limitations of prominent EMR-based predictors that currently affect clinical translation due to significant inter-institutional variability [40].

In this study, the key features selected for the model were derived from time-domain and frequency-domain methods, which uniquely accessible via continuous time series analysis. These features contributed to the model's high predictive

performance. From oxygen saturation, key features extracted included the approximate entropy, permutation entropy, harmonic mean, q statistics, and mode absolute deviation within the 24 hours preceding extubation. These results are highly consistent with the indicators of ROP and intermittent hypoxia previously demonstrated by Di Fiore, et al. [136].

In NICUs, extubation decisions for preterm infants are primarily determined by clinical judgment, resulting in substantial variability in practice and frequent extubation failure [129, 137-139]. While outcomes vary across studies, only 60% to 73% of extremely low birth weight infants are reported to be successfully extubated [138]. Preterm infants who experience extubation failure are at increased risk of respiratory deterioration and fluctuations in cerebral blood flow and oxygenation. Such failure, followed by reintubation, is associated with an extended duration of mechanical ventilation, typically by 10 to 12 days [137-140]. Prolonged use of mechanical ventilation has been associated with an increased incidence of BPD and neurodevelopmental complications [137, 138, 141]. Reintubation, in a select subset of preterm infants, has been shown to increase the risk of BPD or death independently of the length of time spent on mechanical ventilation [139]. Overall, determining the optimal timing for extubation is essential to enhancing both short- and long-term outcomes in preterm infants. While several predictive tools have been developed to assess extubation readiness, consistent and reliable methods remain limited in clinical settings.

This study identified features from real-time patient monitor data to assess extubation readiness, and the resulting predictive model demonstrated high classification accuracy for extubation success. Our findings suggest the critical role of subtle differences in pulmonary oxygenation capacity for successful extubation in preterm infants. We expect these discoveries will form the basis for future high-

performance extubation readiness models, providing clinicians with improved decision-making tools. Furthermore, we anticipate that integrating the key predictors identified in this research with features from prior studies will lead to substantial model improvements and contribute to defining more precise extubation decision guidelines.

※ This chapter is based on the previously published paper, [131] W. Song, Y. Hwa Jung, J. Cho, H. Baek, C. Won Choi, and S. Yoo, "Development and validation of a prediction model for evaluating extubation readiness in preterm infants," *Int J Med Inform*, vol. 178, p. 105192, Oct 2023.

Chapter 5. New physiological Risk Factor of Intraventricular Hemorrhage of Preterm Infant

5.1. Introduction

IVH is a major cause of morbidity in very low birth weight (VLBW) infants and is associated with both short-term [142, 143] and long-term neurodevelopmental impairment [144-146]. The incidence of IVH in VLBW infants was estimated at 50% in the 1970s [146, 147]. With improvements in neonatal intensive care practices, the incidence declined to around 20% by the 1990s [145, 148]. Since the early 2000s, however, IVH rates have remained relatively stable [9, 145, 149]. The pathogenesis of IVH is multifactorial, primarily involving structural immaturity of the cerebral vasculature in preterm infants and impaired autoregulation of cerebral blood flow, both of which contribute to the rupture of fragile vessels within the germinal matrix [150, 151]. IVH has been associated with a range of clinical risk factors reflecting its underlying pathophysiology, including perinatal hypoxic-ischemic injury, respiratory distress syndrome, systemic hypotension, metabolic acidosis, hypercapnia, coagulation and platelet dysfunction, hypothermia, and hyperglycemia [150-152]. Current neonatal intensive care practices have focused on mitigating known risk factors to reduce the incidence of IVH. However, the incidence has not significantly declined and remains high among infants born at <32 weeks' GA or with birth weight <1,500 g [144].

Cranial ultrasound (cUS) is widely used as the standard imaging modality for diagnosing IVH in preterm infants [153]. However, because it is typically performed

at scheduled intervals and IVH often occurs without any clinical signs, the condition may go undetected for several hours or even days after onset [154]. Early identification of infants at risk allows for timely, targeted interventions to reduce the likelihood of further brain injury and improve long-term outcomes [155].

In this study, we identified risk factors for early IVH detection using the previously proposed framework. We also demonstrated the utility of applying analytical methods from other domains, not traditionally used in time series analysis, within our framework for actual IVH identification.

5.2. Methods

5.2.1. Study Design

This study was approved by the Institutional Review Board of SNUBH (IRB No. X-2409-926-902). As a retrospective secondary analysis utilizing de-identified medical records, the requirement for informed consent was waived. The study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for case-control studies [156].

In this study, we included preterm infants admitted to the NICU of SNUBH between March 2018 and December 2022. Infants born at a GA of less than 32 weeks or with a birth weight below 1,500 grams were included in this study. Exclusion criteria included death within 24 hours of birth, major congenital anomalies, missing maternal data, or initiation of vital sign monitoring more than 3 hours after birth.

5.2.2. Data sources

Demographic and clinical variables were extracted from the EMR and NICU

discharge reports. These included antenatal and perinatal factors, maternal and delivery details, neonatal resuscitation data, umbilical cord blood gas values, postnatal arterial blood gas measurements, laboratory test results, and respiratory support parameters, including ventilator settings within the first 24 hours of life. Continuous vital sign data, including heart rate, respiratory rate, oxygen saturation, body temperature, and blood pressure, were recorded at 30s intervals using Philips Patient Monitoring systems equipped with standard clinical measurement devices.

5.2.3. Case and Control Definition

Case infants (IVH group) were identified based on a diagnosis of grade II or higher IVH within the first 7 days of life, confirmed by brain ultrasonography or other imaging modalities. The control group (non-IVH group) consisted of infants with no evidence of IVH beyond germinal matrix hemorrhage (GMH) or grade I IVH. Controls were randomly selected at a 2:1 ratio and individually matched to each case by GA (± 1 week) and birth weight (± 300 g). Magnetic resonance imaging findings were excluded from the analysis to ensure consistency in diagnostic criteria.

5.2.4. Covariates

This study aimed to identify clinically relevant risk factors for early intervention and prevention of IVH-related symptoms using routinely available monitoring data. The analysis was restricted to clinical variables and vital signs documented within the first 24 hours of life. Clinical covariates included demographic information, perinatal factors, IVH-related diagnoses, umbilical cord blood gas parameters, and neonatal resuscitation details. Considering that VLBW infants frequently require respiratory

support, we also calculated the Respiratory Severity Score (RSS), ROX index, and the oxygen saturation to fraction of inspired oxygen ($\text{SpO}_2/\text{FiO}_2$ or SF ratio) ratio [132, 133, 157]. To detect subtle variations in vital signs and potential disruptions in cerebral autoregulation, derived features were extracted using time-series analysis methods, time–frequency domain techniques, and decorrelation time analysis.

5.2.5. Statistical Analysis

Baseline characteristics were described using descriptive statistics. The distribution of continuous variables was assessed for normality using Kolmogorov–Smirnov test. Variables with a normal distribution are reported as mean with standard deviation (SD) and compared using two-tailed Student’s *t*-tests. Non-normally distributed variables were reported as median with interquartile range (IQR) and analyzed using the Mann–Whitney *U* test. Categorical data were compared using either the Chi-Square test or Fisher’s exact test, as appropriate. Variables with more than 50% missing data were excluded from the analysis.

To identify indicators associated with IVH, we applied the continuous vital sign analysis framework developed in this study to select key covariates. We then used multivariable logistic regression, adjusting for GA and birth weight, to assess the associations between IVH outcomes and these candidate covariates. Odds ratios (ORs) were calculated to quantify these associations. ORs and corresponding 95% confidence intervals (CIs) were estimated.

5.3. Results

5.3.1. Study Design

During the study period, 456 infants who met the inclusion criteria were admitted to the NICU. Of these, 70 were excluded due to congenital anomalies (n=7), death within the first day of life (n=4), or insufficient clinical data (n=49). Among the 386 eligible infants, 71 were diagnosed with IVH, stratified by severity as follows: GMH or Grade I (n=42, 59%), Grade II (n=16, 23%), Grade III (n=5, 7%), and Grade IV (n=8, 11%). For the primary analysis, 29 infants with IVH and matched 58 non-IVH controls were selected (Figure 5-1).

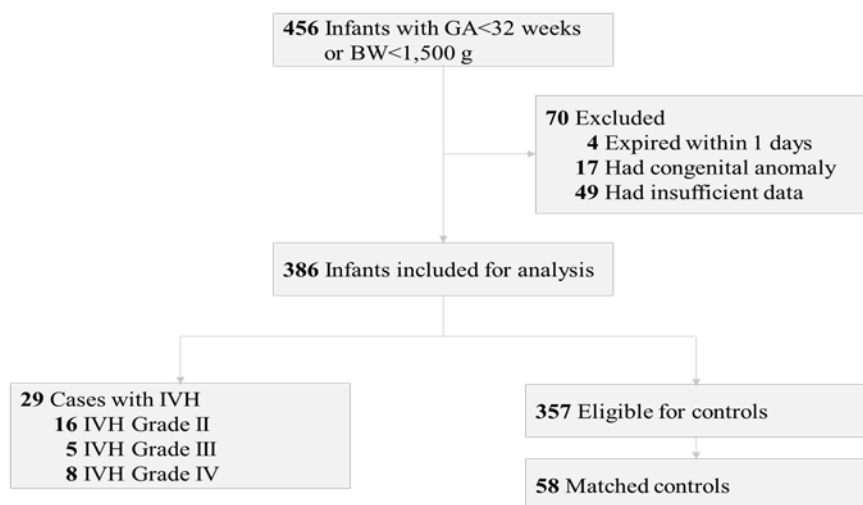


Figure 5-1. Flow diagram of inclusions and exclusions for the study.

Table 5-1 shows a summary of the clinical characteristics of the two groups. The mean (SD) GA was 26.2 (2.7) and 26.9 (2.5) weeks in the IVH and the non-IVH groups, respectively. The mean (SD) birth weight was 865.0g (301.1) and 890.9g (323.0), respectively, with 4 (14.3%) and 14 (24.6%) infants being small for

gestational age (SGA). More infants in the IVH group received invasive ventilation at birth (27 [93.1%]) than those in the non-IVH group (41 [70.7%]). In terms of GA, birth weight, APGAR scores at 1 and 5 min, or prenatal characteristics, the groups were not significantly different. However, the base excess of cord blood gas analysis was less in the IVH group (mean [SD], -5.4 [5.1] mmol/L) than in their matched controls (-3.0 [3.1] mmol/L). No difference was found in the occurrence of persistent pulmonary hypertension among newborns (PPHN). Most clinical characteristics did not significantly differ between the two groups. The median (IQR) time of vital sign measurements after birth was 15 (9.25) min. The time to start patient monitoring did not differ between the IVH and non-IVH groups.

Table 5-1. Baseline characteristics.

	All	Case	Control	p-value
Number of infants	87	29	58	
Perinatal factor				
Maternal age, mean (SD), y	34.7 (3.5)	34.3 (4.1)	34.9 (3.2)	0.445
Maternal underlying disease				
Chorioamnionitis, n (%)	26 (29.9)	10 (34.5)	16 (27.6)	0.679
GDM, n (%)	8 (7.3)	4 (13.8)	3 (5.2)	0.215 [†]
IVF, n (%)	18 (20.7)	6 (20.7)	12 (20.7)	1.000
Oligohydramnios, n (%)	13 (14.9)	1 (3.4)	12 (20.7)	0.052 [†]
Preeclampsia, n (%)	24 (27.6)	4 (13.8)	20 (34.5)	0.075
PROM, n (%)	34 (39.1)	13 (44.8)	21 (36.2)	0.587
Prenatal antibiotics, n (%)	34 (39.1)	11 (37.9)	23 (39.7)	1.000
Antenatal steroid, n (%)	4 (4.6)	3 (10.3)	1 (1.7)	0.106 [†]
Delivery mode, n (%)				
Cesarean section	62 (71.3)	17 (58.6)	45 (77.6)	0.112
Gender, n (%)				
Female	42 (48.3)	12 (41.4)	30 (51.7)	0.495
Male	45 (51.7)	17 (58.6)	28 (48.3)	
Multiple Birth, n (%)	35 (40.2)	14 (48.3)	21 (36.2)	0.395
Gestational age, mean (SD), weeks	26.6 (2.6)	26.2 (2.7)	26.9 (2.5)	0.253
Birth weight, mean (SD), g	882.2 (314.4)	865.0 (301.1)	890.9 (323.0)	0.714
Birth length, mean (SD), cm	34.3 (4.1)	34.5 (4.1)	34.2 (4.1)	0.790
Birth Head Circumference, mean (SD), cm	23.8 (2.9)	24.3 (2.7)	23.7 (3.0)	0.378
APGAR (1 minute), median (IQR)	4.0 [3.0-5.0]	3.0 [2.0-5.0]	4.0 [3.0-5.0]	0.138
APGAR (5 minutes), median (IQR)	7.0 [6.0-8.0]	7.0 [5.0-8.0]	7.0 [6.0-8.0]	0.424
APGAR (1 minute)<7, n (%)	81 (93.1)	26 (89.7)	55 (94.8)	0.396
APGAR (5 minutes)<7, n (%)	30 (34.5)	12 (41.4)	18 (31.0)	0.473
SGA, n (%)	18 (21.2)	4 (14.3)	14 (24.6)	0.419 [†]
Cord blood gas analysis				
pH, mean (SD)	7.3 (0.1)	7.3 (0.1)	7.3 (0.1)	0.138
BE, mean (SD), mmol/L	-3.8 (4.0)	-5.4 (5.1)	-3.0 (3.1)	0.045
PCO ₂ , mean (SD), mmHg	48.1 (13.7)	50.8 (19.7)	46.7 (9.6)	0.342
Resuscitation				
PPV, n (%)	67 (77.0)	19 (65.6)	48 (82.8)	0.126
Intubation, n (%)	58 (66.7)	21 (72.4)	37 (63.8)	0.574
Epinephrine, n (%)	2 (2.3)	1 (3.4)	1 (1.7)	1.000 [†]
CM, n (%)	2 (2.3)	1 (3.4)	1 (1.7)	1.000 [†]
RDS, n (%)	73 (83.9)	26 (89.7)	47 (81.0)	0.369
Ventilatory support mode within 24 hours after birth				
Invasive Ventilation, n (%)	68 (78.2)	27 (93.1)	41 (70.7)	0.035
HFOV, n (%)	27 (39.7)	11 (40.7)	16 (39.0)	1.000
Conventional Ventilation, n (%)	41 (60.3)	16 (59.3)	25 (61.0)	1.000
Inhaled NO within 24 hours after birth, n (%)	14 (16.1)	8 (27.6)	6 (10.3)	0.061
Inotropics administration within 24 hours, n (%)	1 (1.1)	0 (0.0)	1 (1.7)	1.000 [†]
Laboratory finding within 24 hours after birth				
Blood gas analysis				
PCO ₂ , mean (SD), mmHg	42.3 (7.3)	43.0 (6.6)	42.0 (7.6)	0.524
pH, mean (SD)	7.3 (0.1)	7.2 (0.1)	7.3 (0.1)	0.112
Hemoglobin, mean (SD), g/dl	14.8 (2.1)	14.7 (2.4)	14.8 (2.0)	0.915

Abbreviations: BE, base excess of cord blood gas analysis; CM, cardiac massage; FiO₂, fraction of inspired oxygen; GDM, gestational diabetes mellitus; HFOV, high frequency oscillatory ventilation; IVF, in vitro fertilization; NO, nitric oxide; PCO₂, partial pressure of carbon dioxide; PPV, positive-pressure ventilation; PROM, premature rupture of membranes; RDS, respiratory distress syndrome.

5.3.2. Risk Factors Associated with IVH

Table 5-2 shows the results of univariable and multivariable analyses of the selected covariates. From 247,000 continuous vital sign candidate features, 20 features were selected. Several vital signal-related risk factors met the predefined p-value threshold. Among the demographic variables, we included preeclampsia, base excess, and resuscitation with positive pressure ventilation (PPV). After adjusting for multicollinearity, three covariates remained; only SpO₂ decorrelation time was found to be significantly associated with IVH in the multivariable analysis (Table 5-2). An increase in SpO₂ decorrelation time was associated with a higher risk of IVH (adjusted OR [aOR], 1.53; 95% CI, 1.08–2.17 for per minute increase). Infants with SpO₂ decorrelation time > 5.62 minutes (56.2%), based on the optimal cutoff, had an 11-fold increased risk of IVH compared with infants without such prolonged decorrelation time (aOR, 11.35; 95% CI, 3.54–36.38).

Table 5-2. Univariate and multivariable analysis of risk factors associated with IVH.

Risk Factor	Univariate Association				Multivariable			
	Adjusted Results		Unadjusted Results		Adjusted Results		Unadjusted Results	
	aOR (95% CI)	p-value	OR (95% CI)	p-value	aOR (95% CI)	p-value	OR (95% CI)	p-value
SpO ₂ Decorrelation Time	1.81 (1.31–2.51)	0.0004	1.80 (1.31–2.46)	0.0003	1.58 (1.10–2.25)	0.0124	1.52 (1.07–2.16)	0.0186
SBP, MAD of Δx (0h–12h)	0.46 (0.28–0.74)	0.0013	0.49 (0.32–0.75)	0.0011	0.68 (0.39–1.20)	0.1873	0.76 (0.47–1.21)	0.2448
SBP, Median of Δx (0h–6h)	0.68 (0.55–0.85)	0.0008	0.70 (0.58–0.86)	0.0004	0.85 (0.67–1.08)	0.1741	0.84 (0.68–1.05)	0.1220

Abbreviations: aOR, adjusted odds ratio; CI, confidence interval; MAD, mean absolute deviation; SBP, systolic blood pressure; SpO₂, oxygen saturation.

5.3.3. Oxygen Saturation Decorrelation Time

The mean time for SpO₂ decorrelation and raw SpO₂ levels during the first 24h after birth, and their 95% CIs, are shown in Figure 5-2, Figure 5-3 and Figure 5-4, stratified by IVH and non-IVH groups. During the first 4h after birth, SpO₂ decorrelation time trajectories were highly unstable in both groups. The non-IVH group exhibited a stable trajectory 7h after birth. In contrast, the IVH group showed

significantly longer decorrelation times than the non-IVH group, indicating that the first SpO₂ decorrelation time was sustained for an increased duration in the IVH group. Additionally, SpO₂ decorrelation time in the non-IVH group was significantly lower than those in the IVH group after the 7-hour mark. However, the raw SpO₂ trends were not clearly different between the two groups, with most SpO₂ values remaining >95%.

For a more detailed analysis of the fluctuations related to IVH, we visualized the autocorrelation values for each differential order and highlighted the variations across the time lags using density plots (Figure 5-4). In SpO₂ autocorrelation, which reflects persistent instability in oxygen saturation, the density of autocorrelation values in the non-IVH group spread toward lower values after birth. However, the IVH group maintained consistently high values (Figure 5-4 A and B). For the first- and second-order differentials, which show instability in the rate of SpO₂ changes, the IVH group showed sustained instability in SpO₂ change rates, consistent with autocorrelation spanning 2–4 min (20%–40%). (Figure 5-4 C and D). Conversely, in the non-IVH group, the autocorrelation values decayed to zero 6 h post-birth (Figure 5-4 E and F). Additionally, based on the density plots, infants in the non-IVH group appeared to recover faster from SpO₂ changes and stabilized within a few minutes. In contrast, infants in the IVH group showed longer oxygen instability that persisted for >5 min (Figure 5-2 A and B).

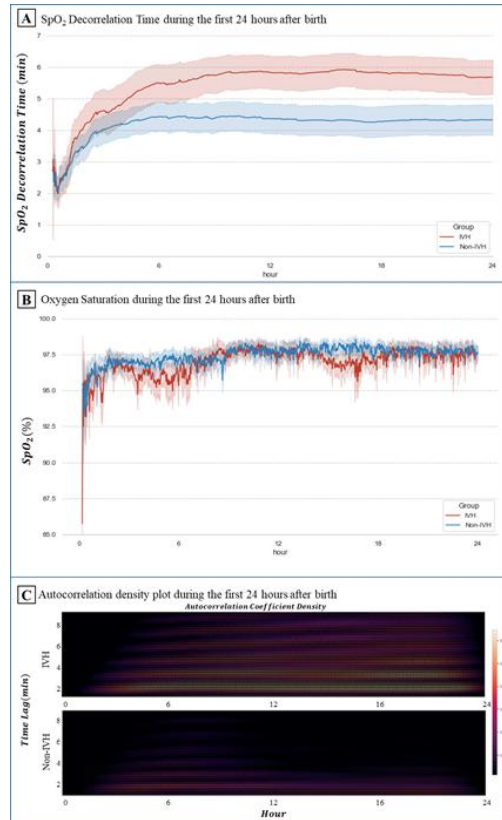


Figure 5-2. Trend and instability density in SpO₂ decorrelation time. (A) Line plot of SpO₂ decorrelation time over the first 24 h after birth in the IVH and non-IVH groups. The shaded area shows 95% CI of the SpO₂ decorrelation time. (B) Line plot of raw SpO₂ over the first 24 h after birth in the IVH and non-IVH groups. The shaded area represents the 95% CI of SpO₂. (C) Density plot of autocorrelation by time lag in the IVH and non-IVH groups. The x-axis represents the elapsed hours after birth, the y-axis represents the time lag, and the color intensity indicates the projections of autocorrelation values by time lag.

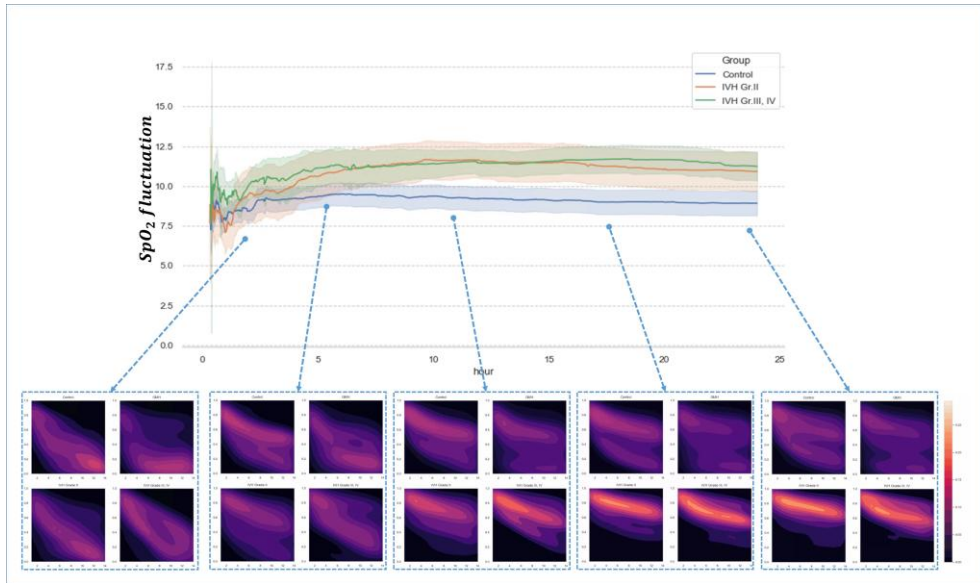


Figure 5-3. Trend and density plot of SpO₂ decorrelation time across groups in the First 24 hours after birth.

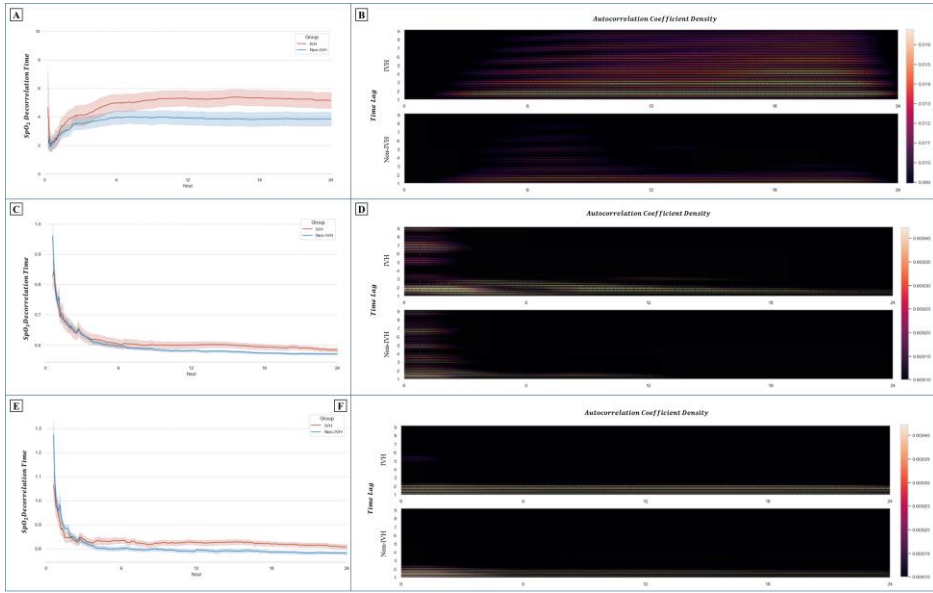


Figure 5-4. Trend and density plot of SpO₂ decorrelation time in each order differential. (A) Line plot of SpO₂ decorrelation time over the first 24 h after birth in the IVH and non-IVH groups. (B) Density plot of autocorrelation by time lag in the IVH and non-IVH groups. (C) Line plot of first-order differential SpO₂ decorrelation time over the first 24 h after birth in the IVH and non-IVH groups. (D) Density plot of first-order differential SpO₂ autocorrelation by time lag in the IVH and non-IVH groups. (E) Line plot of second-order differential SpO₂ decorrelation time over the first 24 h after birth in the IVH and non-IVH groups. (F) Density plot of second-order differential SpO₂ autocorrelation by time lag in the IVH and non-IVH groups.

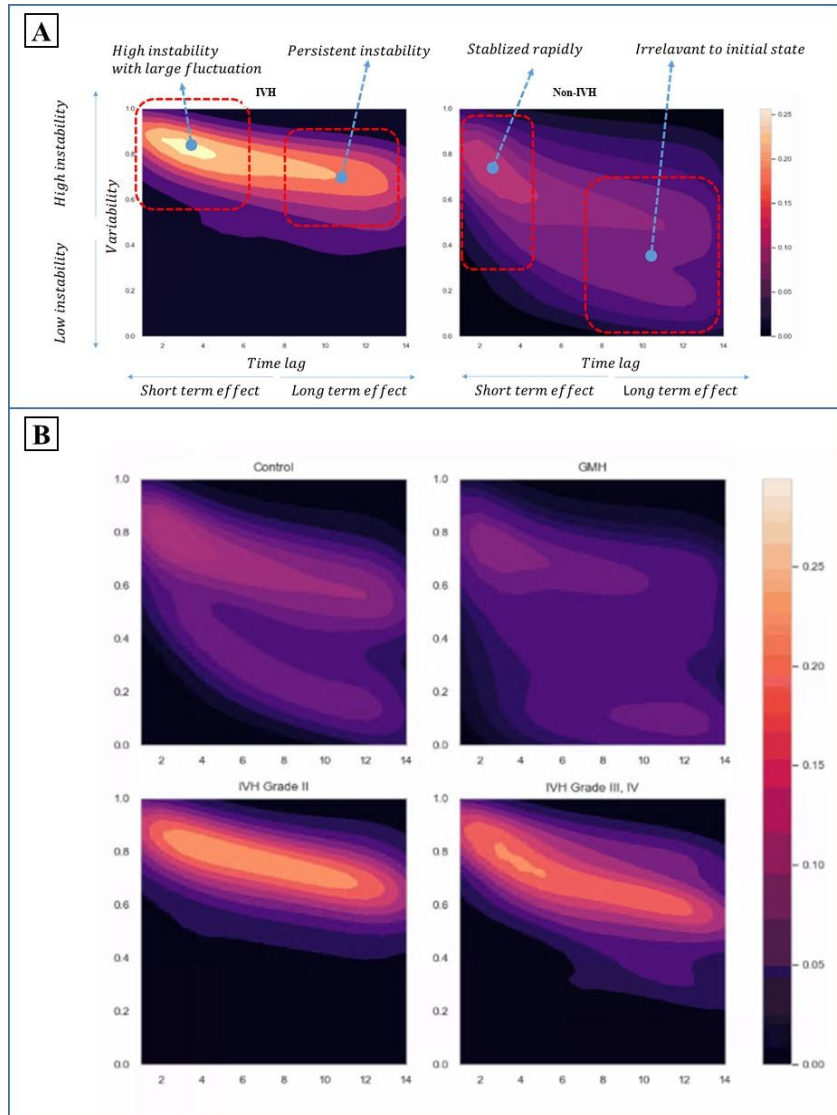


Figure 5-5. Autocorrelation density plot of SpO₂ decorrelation time at 18 hours post-birth. (A) Density plot of autocorrelation in the IVH and non-IVH groups. (B) Density plot of autocorrelation by time lag in the IVH with Grade I, II, III and IV groups.

5.3.4. Sensitivity Analyses

The results of the sensitivity analysis of the minimum sampling rate were obtained by undersampling each epoch (Figure 5-6). SpO₂ decorrelation time in the 1-minute sampling period (1/60 Hz) was similar to that of the primary analysis. However, no significant differences were observed between the two groups for sampling periods longer than 5min sampling period. The results of the sensitivity analyses using the modified case-control definition to examine potential bias due to matching methods or IVH grade were consistent with those of the primary analysis (Table 5-3. Ratios of risk factors for each). The results of the regression model sensitivity analyses, to examine the possible bias due to the regression model, were consistent with those of the primary analysis (Table 5-4).

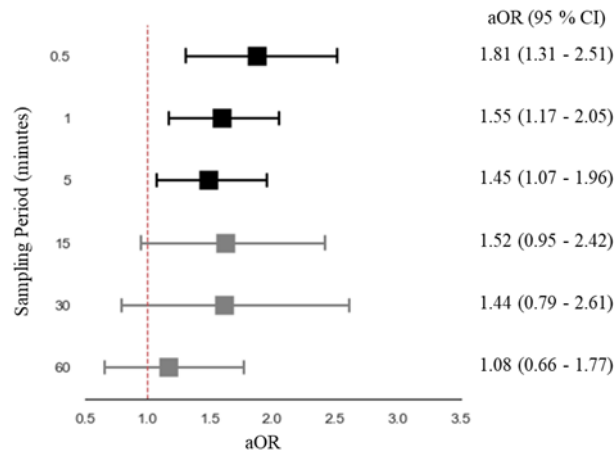


Figure 5-6. Forest plot of SpO₂ decorrelation time for each sampling period.

Table 5-3. Ratios of risk factors for each group.

Risk Factor	Univariate				Multivariate			
	Unadjusted Results OR (95% CI)	p-value	Adjusted Results aOR (95% CI)	p-value	Unadjusted Results OR (95% CI)	p-value	Adjusted Result aOR (95% CI)	p-value
Matched IVH (Any Grade)/Non-IVH								
SpO ₂ Decorrelation	1.47 (1.25-1.73)	<0.0001	1.41 (1.19-1.69)	0.0001	1.42 (1.20-1.68)	<0.0001	1.38 (1.15-1.65)	0.0004
SBP, MAD of Δx (0h-12h)	0.83 (0.68-1.01)	0.0569	0.87 (0.72-1.05)	0.1567	0.97 (0.78-1.19)	0.7595	0.97 (0.79-1.20)	0.8127
SBP, Median of Δx (0h-6h)	0.89 (0.81-0.97)	0.0106	0.91 (0.83-0.99)	0.0302	0.93 (0.84-1.03)	0.1714	0.97 (0.83-1.14)	0.1835
Matched IVH/Control (Non-IVH+GMH)								
SpO ₂ Decorrelation	1.80 (1.34-2.41)	0.0001	1.79 (1.32-2.43)	0.0002	1.51 (1.10-2.08)	0.0113	1.55 (1.12-2.15)	0.0082
SBP, MAD of Δx (0h-12h)	0.50 (0.33-0.73)	0.0004	0.49 (0.33-0.75)	0.0008	0.80 (0.52-1.24)	0.3215	0.75 (0.46-1.23)	0.2579
SBP, Median of Δx (0h-6h)	0.68 (0.56-0.83)	0.0001	0.67 (0.54-0.83)	0.0003	0.81 (0.65-1.00)	0.0549	0.86 (0.62-1.20)	0.0803
Unmatched IVH/Control								
SpO ₂ Decorrelation Time	2.40 (1.80-3.18)	<0.0001	2.09 (1.53-2.86)	<0.0001	2.05 (1.51-2.79)	<0.0001	1.93 (1.39-2.66)	0.0001
SBP, MAD of Δx (0h-12h)	0.35 (0.23-0.54)	<0.0001	0.50 (0.33-0.77)	0.0013	0.55 (0.32-0.94)	0.0285	0.65 (0.39-1.10)	0.1104
SBP, Median of Δx (0h-6h)	0.58 (0.47-0.73)	<0.0001	0.71 (0.58-0.87)	0.0011	0.84 (0.67-1.05)	0.1228	0.87 (0.70-1.08)	0.2219
Unmatched IVH (Any Grade)/Control								
SpO ₂ Decorrelation Time	1.52 (1.30-1.78)	<0.0001	1.41 (1.19-1.67)	0.0001	1.48 (1.26-1.74)	<0.0001	1.38 (1.16-1.64)	0.0003
SBP, MAD of Δx (0h-12h)	0.79 (0.65-0.96)	0.0185	0.87 (0.72-1.04)	0.1299	0.92 (0.74-1.14)	0.4313	0.95 (0.77-1.17)	0.6136
SBP, Median of Δx (0h-6h)	0.89 (0.81-0.97)	0.0082	0.92 (0.84-0.99)	0.0455	0.95 (0.86-1.05)	0.3005	0.95 (0.86-1.05)	0.3250
Unmatched IVH/Control (Non-IVH+GMH)								
SpO ₂ Decorrelation Time	2.31 (1.75-3.05)	<0.0001	1.97 (1.45-2.67)	<0.0001	1.95 (1.45-2.62)	<0.0001	1.81 (1.33-2.48)	0.0002
SBP, MAD of Δx (0h-12h)	0.36 (0.24-0.55)	<0.0001	0.52 (0.34-0.78)	0.0014	0.62 (0.37-1.02)	0.0621	0.71 (0.44-1.16)	0.1744
SBP, Median of Δx (0h-6h)	0.58 (0.46-0.72)	<0.0001	0.71 (0.58-0.87)	0.0009	0.82 (0.65-1.03)	0.0830	0.85 (0.69-1.05)	0.1287

Abbreviations: aOR, adjusted odd ratio; MAD, median absolute deviation; SBP, systolic blood pressure.

Table 5-4. Odds ratios of risk factors for various regression models.

Risk Factor	LR		LASSO		GLM	
	aOR (95%CI)	p-value	aOR (95%CI)	p-value	aOR (95%CI)	p-value
SpO ₂ fluctuation	1.58 (1.10-2.25)	0.0124	1.57 (1.11-2.21)	0.0109	1.09 (1.03-1.15)	0.0020
SBP, MAD of Δx (0h-12h)	0.68 (0.39-1.20)	0.1873	0.68 (0.39-1.21)	0.1892	0.94 (0.89-1.00)	0.0522
SBP, Median of Δx (0h-6h)	0.85 (0.67-1.08)	0.1741	0.85 (0.67-1.08)	0.1715	0.98 (0.96-1.01)	0.1936

Abbreviations: aOR, adjusted odd ratio; MAD, median absolute deviation; MBP, mean blood pressure; SBP, systolic blood pressure; LR, logistic regression; LASSO, least absolute shrinkage and selection operator; GLM, generalized linear model.

5.4. Discussion

In this study, preterm infants who developed IVH exhibited significantly greater oxygen saturation variability and prolonged periods of instability, as quantified by SpO₂ decorrelation time, during the first 24 hours of life compared with infants without IVH.

Decorrelation time, a metric widely used in non-medical domains to quantify regional persistence or variations in decay time, gives unique advantages for analyzing time-series data [158-163]. Previous studies showed that decorrelation

time is affected by fluctuations or variations with amplitudes exceeding those of white noise. Specifically, slower decay rates and sustained oscillatory variations are linked to prolonged decorrelation times. Notably, the decorrelation time obtained from biosignals with varying amplitudes and heterogeneous patterns within the same observation window is useful for inferring delayed autoregulatory responses in specific signal components. Theoretically, under the weak stationary assumption, contributing white noise to decorrelation time converges to zero, making it a robust metric of physiological variability. Therefore, only SpO₂ met these conditions and was seen as a reliable risk factor for IVH. Considering these attributes, decorrelation time was hypothesized to provide a reliable indicator of recovery time and autoregulatory function in neonates, offering a novel tool for assessing physiological stability in this vulnerable population.

Mean and raw continuous SpO₂ levels within the first 24 hours of life were similar between the IVH and non-IVH groups, suggesting that oxygen saturation fluctuations may be subtle and not readily detectable in routine clinical assessment. Additionally, no significant differences were observed in respiratory support parameters, including FiO₂, RSS, and ROX index, between the groups. These findings reflect intrinsic differences in respiratory physiology rather than variations in clinician-directed management.

The regulation of cerebral blood flow during hypoxemia involves multiple integrated mechanisms that promote vasodilation. These include direct action on vascular smooth muscle, endothelium-mediated pathways, and the release of signaling molecules such as adenosine and potassium ions from neurons and glial cells. Furthermore, the hypoperfusion-reperfusion cycle that can result from such vascular changes is a key factor in the pathogenesis of IVH in preterm infants [37, 164-169].

Given that minor SpO₂ fluctuations can correspond to major PaO₂ changes in preterm infants, persistent instability may trigger damaging cerebral hypoperfusion-reperfusion cycles in the vulnerable germinal matrix [37, 150]. In this study, we observed a critical divergence in SpO₂ patterns between infants who did and did not develop IVH, occurring after the initial phase of postnatal adaptation [170, 171]. While both groups had unstable SpO₂ during the first four hours, infants in the non-IVH group achieved stability by seven hours. In contrast, infants who developed IVH demonstrated sustained fluctuations and a longer recovery from hypoxic episodes. This suggests that the inability to stabilize SpO₂ after the first six hours of life, rather than early fluctuation itself, may be a key early marker of impaired cerebral autoregulation and heightened IVH risk.

Previous studies have identified key risk factors for IVH, leading to established prevention and management strategies. Perinatal interventions primarily involve preventing premature birth, optimizing labor and delivery (e.g., antenatal glucocorticoids, delayed cord clamping, thermal stability), and providing high-quality respiratory care [172-175]. Postnatal efforts aim to stabilize cerebral blood flow through nursing bundles, slow blood draws, correction of hemodynamic and coagulation abnormalities, and pharmacological therapies such as phenobarbital and indomethacin [175-177]. However, persistent challenges in effectively stabilizing cerebral blood flow hinder further reductions in IVH occurrence.

While studies have attempted to predict IVH occurrence, a clinically applicable model has yet to be developed [155, 169, 178-180]. A key impediment is the inability to confirm IVH onset in real time. Standard bedside cranial ultrasound, being a manually conducted procedure, complicates accurate temporal diagnosis. However, given that most IVH cases manifest within 72 hours of birth, with approximately 50% emerging within the first 24 hours, hemodynamic fluctuations during this initial 24-

hour window may be closely linked to its development [154, 180-182].

Recent research has explored early markers for IVH in extremely preterm infants. Iyer, et al. [155] quantitatively assessed electroencephalography (EEG) during the first 72 hours of life in 25 infants, identifying sharper and less symmetric EEG burst shapes as early indicators of IVH. Cimatti, et al. [169] investigated changes in cerebral oxygenation (CrSO_2), cerebral fractional oxygen extraction (cFTOE), and the tissue oxygenation-heart rate reactivity index (TOHRx) preceding and following IVH occurrence within the same 72-hour postnatal period. In infants who developed IVH, CrSO_2 demonstrated an initial increase followed by a plateau, while cFTOE decreased before subsequently rising, with peak changes occurring between 24 and 48 hours. Conversely, these indicators remained stable in infants without IVH, underscoring the role of impaired cerebral autoregulation in IVH pathogenesis. These novel bedside measures exhibit high diagnostic accuracy, potentially enabling IVH detection prior to ultrasound confirmation, thus offering opportunities for earlier intervention and personalized care. This study similarly investigates differences between groups using real-time neonatal monitoring data. However, current real-time monitoring methods like EEG and near-infrared spectroscopy (NIRS) have limitations. EEG requires specialized equipment and expertise for interpretation and is prone to artifacts. NIRS, while more accessible than EEG, is susceptible to detection errors from external light, skin thickness, and movement artifacts. Furthermore, signal fluctuations and inherent variability in infant cerebral oxygenation hinder the establishment of universal predictive thresholds for IVH. Consequently, these methods face limitations compared to more accessible and widely used SpO_2 monitoring.

To enable precise monitoring of subtle vital sign changes, we utilized real-time vital sign data, recorded every 30 seconds, from the first 24 hours of life. Our

comprehensive analysis, employing descriptive statistics, time series analysis, time-frequency domain analysis, and autocorrelation methods, revealed that differences in SpO₂ decorrelation times were detectable at shorter intervals (1–5 minutes). These differences were indistinguishable with data recorded at intervals of <15 minutes, underscoring the limitation of traditional medical records that rely on longer measurement intervals for identifying subtle, clinically significant variations.

A primary strength of this study lies in its capacity to detect subtle, clinically imperceptible variations in SpO₂ through real-time vital sign monitoring at 30-second intervals, even when raw SpO₂ values consistently remain above 95%. Our analysis of these real-time physiological indicators offers a novel method for determining the probability of IVH occurrence, thereby facilitating bedside clinical decision-making and enabling more precise respiratory and cardiovascular management of preterm infants in the NICUs.

Our study was limited by its retrospective design. First, we could not include direct measures of cerebral perfusion/oxygenation (e.g., NIRS) or comprehensive cardiac function via echocardiography. We also lacked real-time values for invasive blood gas parameters, which are known to significantly influence cerebral blood flow. However, we demonstrated that the frequency and mean values of blood gas analyses did not differ between the IVH and non-IVH groups during the first 24 hours of life. Future research should address these gaps by incorporating non-invasive real-time monitoring methods, such as transcutaneous CO₂ monitoring and NIRS.

※ This chapter will be submitted to a peer-reviewed journal for publication.

Chapter 6. Conclusion

This study introduces an efficient and scalable methodology and framework for the extraction and selection of features from continuous vital signs, an area holding substantial recent promise. This methodology was specifically designed to address the unique characteristics of vulnerable preterm infants and their continuous vital signs in the NICU. Furthermore, we investigated clinically relevant risk factors and developed a predictive model, thereby substantiating the feasibility and applicability of continuous vital signs through external validation.

Our methodology features a flexible pipeline, enabling easy integration of diverse feature calculation methods and direct analysis of their clinical significance. We further applied case-control emulation and FDR control method using established clinical statistical tests and estimators. This approach mitigates false positives from multiple comparisons, ensuring reliable results. Notably, the process was designed to be partitionable, enhancing scalability.

To validate our methodology, we implemented the framework utilizing a distributed computing architecture based on the MapReduce model's divide-and-conquer concept. This implementation was feasible due to the unidirectional design of our analysis methods and their reliance on separate resample or subsample-based analysis, maintaining reliability during computation. Although our current implementation operated on a single server and used a basic MapReduce model, it has the potential for future extension to include real-time analysis capabilities through features like in-memory databases or optimized NoSQL-based aggregation, which will be explored in future research. We anticipate that this validated and robust methodology can be expanded to integrate data from various institutions or by incorporating additional deep learning nodes.

The features derived from our proposed methods exhibited distinct characteristics for each morbidity. In sepsis and all-cause mortality study, we observed a negative correlation between sepsis and heart rate entropy, consistent with previous HRC research. This implied that heart rate variability and entropy, often collectable from patient monitor at 0.5–1 Hz in most NICUs, can achieve similar performance to ECG-based systems like the HeRO score. The inclusion of pulse oximeter entropy with similar or identical contributions to heart rate features suggests a promising for detecting infection and deterioration in patients for whom ECG measurement is challenging, or for discharged neonates in home care settings, given the ease of use of pulse oximeters. In extubation readiness, while previous extubation readiness predictive models showed valid performance even across different internal validation timeframes, we identified that integrating continuous vital signs significantly enhances the accuracy of patient status assessment. In IVH study, our methodology successfully identified SpO₂ instability, derived using the decorrelation time method from outside the traditional medical domain, as a novel physiological marker for IVH onset detection. We expect that this newly identified physiological marker could enable proactive interventions before severe IVH develops, thereby improving patient outcomes.

This study has limitations across technical, statistical, and clinical domains, which inform areas for future research.

From a technical perspective, the limitations are as follows. Firstly, the MapReduce model utilized in this study does not represent the current state-of-the-art methodology. Consequently, the implemented framework may exhibit significantly slower processing performance and higher latency compared to contemporary approaches, such as those employing in-memory databases. However, since the algorithms and methods proposed in this study are fundamentally

compatible with distributed computing, they can be readily adapted to more advanced database technologies or those supporting efficient aggregation. Therefore, future research should investigate which database and architectural configurations are most efficient for identifying key risk factors and features across various database systems.

Secondly, our proposed framework presents security vulnerabilities. Due to scalability concerns with CouchDB's default JavaScript-based query server, this study developed a custom Python-based query server. However, systems that execute scripts, such as Python, are susceptible to exploits from external intrusions. This inherent risk has led to a recent trend where external script-supporting query servers are only offered with limited functionality. Accordingly, enhancing the security of data transmission and the query server itself is crucial for the practical application of this framework.

Thirdly, this study did not implement specific load balancing or data redistribution mechanisms. A bottleneck in a single server can consequently extend the overall execution time. Therefore, any real-world application of this framework would necessitate incorporating strategies such as task replication and data redundancy to ensure robust performance.

Finally, the reported performance metrics were derived from a Kubernetes cluster simulated using Kind. As such, these simulations do not account for network latency or bandwidth, nor do they involve physically distinct servers. Consequently, the actual performance when deploying the framework across multiple physically separate nodes may differ considerably. Therefore, future research should further evaluate the impact of network considerations on performance.

From a statistical perspective, the limitations are as follows. First, further research is required to determine the optimal number of resampling iterations and an

appropriate cutoff value. This study applied a minimum of 200 resampling iterations; additional analysis is needed to assess the extent to which a greater number of iterations or resampling improves statistical power and FDR mitigation.

Second, the estimators used in this study do not reflect modern approaches. Recent research, for example, has explored applying deep learning-based estimators, such as the X-model or knockoff filter. Consequently, this study did not investigate the specific characteristics or performance implications associated with different types of estimators. Therefore, further research is needed to investigate the characteristics of each estimator.

The clinical limitations are as follows. First, data scarcity for external validation posed a significant challenge. Our analyses exclusively utilized the UVA NICU dataset for external validation. This dataset's focus on all-cause mortality also restricted our capacity to conduct direct, event-specific performance comparisons. These constraints stem from the general scarcity of continuous vital sign databases for NICU preterm infants, coupled with a lack of linked demographic or diagnosis data in existing repositories.

Second, the limited subject numbers within specific event cohorts restricted the study scope. The count of preterm infants included in certain event analyses was notably restricted. For instance, the IVH cohort comprised only 29 infants. Consequently, further research is crucial to ascertain the external validity of the identified predictors across diverse institutional settings.

This study overcomes limitations of computational burden and restricted time series analysis in conventional clinical research. By advancing continuous vital sign research and its clinical utility, our work aims to improve research efficiency and address clinical needs.

Bibliography

- [1] R. E. Behrman and A. S. Butler, *Preterm birth : causes, consequences, and prevention*. National Academies Press, 2007, pp. 772–772.
- [2] A. Seassau, P. Munos, C. Gire, B. Tosello, and I. Carchon, "Neonatal Care Unit Interventions on Preterm Development," *Children (Basel)*, vol. 10, no. 6, Jun 2 2023, doi: 10.3390/children10060999.
- [3] J. V. Gill and E. M. Boyle, "Outcomes of infants born near term," in *Archives of Disease in Childhood* vol. 102, ed: BMJ Publishing Group, 2017, pp. 194–198.
- [4] A. S. Morgan, M. Mendonça, N. Thiele, and A. L. David, "Management and outcomes of extreme preterm birth," *The BMJ*, vol. 376, 2022/1// 2022, doi: 10.1136/bmj-2021-055924.
- [5] D. M. Ely and A. K. Driscoll, "National Vital Statistics Reports Volume 73, Number 5 July 25,2024 Infant Mortality in the United States, 2022: Data From the Period Linked Birth/Infant Death File," in "National Vital Statistics Reports," 2024, vol. 73. [Online]. Available: <https://www.cdc.gov/nchs/products/index.htm>.
- [6] T. Venkatesan *et al.*, "National Trends in Preterm Infant Mortality in the United States by Race and Socioeconomic Status, 1995-2020," *JAMA Pediatrics*, vol. 177, no. 10, pp. 1085–1095, 2023/10// 2023, doi: 10.1001/jamapediatrics.2023.3487.
- [7] E. O. Ohuma *et al.*, "National, regional, and global estimates of preterm birth in 2020, with trends from 2010: a systematic analysis," *The Lancet*, vol. 402, no. 10409, pp. 1261–1271, 2023/10// 2023, doi: 10.1016/S0140-6736(23)00878-4.
- [8] J. Perin *et al.*, "Global, regional, and national causes of under-5 mortality in 2000–19: an updated systematic analysis with implications for the Sustainable Development Goals," *The Lancet Child and Adolescent Health*, vol. 6, no. 2, pp. 106–115, 2022/2// 2022, doi: 10.1016/S2352-4642(21)00311-4.
- [9] E. F. Bell *et al.*, "Mortality, In-Hospital Morbidity, Care Practices, and 2-Year Outcomes for Extremely Preterm Infants in the US, 2013-2018," *JAMA*, vol. 327, no. 3, pp. 248–263, 2022, doi: 10.1001/jama.2021.23580.
- [10] C. P. Hornik *et al.*, "Early and late onset sepsis in very-low-birth-weight infants from a large group of neonatal intensive care units," *Early Human Development*, vol. 88, pp. S69–S74, 2012/05/01/ 2012, doi: [https://doi.org/10.1016/S0378-3782\(12\)70019-1](https://doi.org/10.1016/S0378-3782(12)70019-1).
- [11] B. J. Stoll *et al.*, "Late-onset sepsis in very low birth weight neonates: A report from the National Institute of Child Health and Human Development Neonatal Research Network," *The Journal of Pediatrics*, vol. 129, no. 1, pp. 63–71, 1996, doi: 10.1016/S0022-3476(96)70191-9.
- [12] S. J. Schrag and J. R. Verani, "Intrapartum antibiotic prophylaxis for the prevention of perinatal group B streptococcal disease: Experience in the United States and implications for a potential group B streptococcal vaccine," *Vaccine*, vol. 31, pp. D20–D26, 2013/08/28/ 2013, doi: <https://doi.org/10.1016/j.vaccine.2012.11.056>.
- [13] H. Patural, V. Pichot, F. Roche, and A. Giraud, "Why, when and how to assess autonomic nervous system maturation in neonatal care units: A practical overview," *Neurophysiologie Clinique*, vol. 53, no. 2, p. 102855, 2023/04/01/ 2023, doi: <https://doi.org/10.1016/j.neucli.2023.102855>.
- [14] R. Tiwari, R. Kumar, S. Malik, T. Raj, and P. Kumar, "Analysis of Heart Rate Variability and Implication of Different Factors on Heart Rate Variability," *Current Cardiology Reviews*, vol. 17, no. 5, pp. 74–83, 2021, doi:

<https://doi.org/10.2174/1573403X16999201231203854>.

- [15] M. Malik *et al.*, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *European Heart Journal*, vol. 17, no. 3, pp. 354–381, 1996, doi: 10.1093/oxfordjournals.eurheartj.a014868.
- [16] K. D. Fairchild and T. M. O'Shea, "Heart Rate Characteristics: Physiomarkers for Detection of Late-Onset Neonatal Sepsis," *Clinics in Perinatology*, vol. 37, no. 3, pp. 581–598, 2010/09/01/ 2010, doi: <https://doi.org/10.1016/j.clp.2010.06.002>.
- [17] M. L. Stone *et al.*, "Abnormal heart rate characteristics before clinical diagnosis of necrotizing enterocolitis," *Journal of Perinatology*, vol. 33, no. 11, pp. 847–850, 2013/11/01 2013, doi: 10.1038/jp.2013.63.
- [18] M. P. Griffin, T. M. O'Shea, E. A. Bissonette, F. E. Harrell, D. E. Lake, and J. R. Moorman, "Abnormal Heart Rate Characteristics Preceding Neonatal Sepsis and Sepsis-Like Illness," *Pediatric Research*, vol. 53, no. 6, pp. 920–926, 2003/06/01 2003, doi: 10.1203/01.PDR.0000064904.05313.D2.
- [19] M. P. Griffin and J. R. Moorman, "Toward the Early Diagnosis of Neonatal Sepsis and Sepsis-Like Illness Using Novel Heart Rate Analysis," *Pediatrics*, vol. 107, no. 1, pp. 97–104, 2001, doi: 10.1542/peds.107.1.97.
- [20] J. R. Moorman *et al.*, "Mortality Reduction by Heart Rate Characteristic Monitoring in Very Low Birth Weight Neonates: A Randomized Trial," *The Journal of Pediatrics*, vol. 159, no. 6, pp. 900–906.e1, 2011, doi: 10.1016/j.jpeds.2011.06.044.
- [21] K. Addison, M. P. Griffin, J. R. Moorman, D. E. Lake, and T. M. O'Shea, "Heart rate characteristics and neurodevelopmental outcome in very low birth weight infants," *Journal of Perinatology*, vol. 29, no. 11, pp. 750–756, 2009/11/01 2009, doi: 10.1038/jp.2009.81.
- [22] B. A. Sullivan, C. McClure, J. Hicks, D. E. Lake, J. R. Moorman, and K. D. Fairchild, "Early Heart Rate Characteristics Predict Death and Morbidities in Preterm Infants," *The Journal of Pediatrics*, vol. 174, pp. 57–62, 2016, doi: 10.1016/j.jpeds.2016.03.042.
- [23] K. K. Doheny *et al.*, "Diminished vagal tone is a predictive biomarker of necrotizing enterocolitis-risk in preterm infants," *Neurogastroenterology & Motility*, vol. 26, no. 6, pp. 832–840, 2014/06/01 2014, doi: <https://doi.org/10.1111/nmo.12337>.
- [24] K. D. Fairchild *et al.*, "Abnormal heart rate characteristics are associated with abnormal neuroimaging and outcomes in extremely low birth weight infants," *Journal of Perinatology*, vol. 34, no. 5, pp. 375–379, 2014/05/01 2014, doi: 10.1038/jp.2014.18.
- [25] N. Goel, M. Chakraborty, W. J. Watkins, and S. Banerjee, "Predicting Extubation Outcomes: A Model Incorporating Heart Rate Characteristics Index," *The Journal of Pediatrics*, vol. 195, pp. 53–58.e1, 2018, doi: 10.1016/j.jpeds.2017.11.037.
- [26] B. A. Sullivan, S. M. Grice, D. E. Lake, J. R. Moorman, and K. D. Fairchild, "Infection and Other Clinical Correlates of Abnormal Heart Rate Characteristics in Preterm Infants," *The Journal of Pediatrics*, vol. 164, no. 4, pp. 775–780, 2014, doi: 10.1016/j.jpeds.2013.11.038.
- [27] N. Kumar, G. Akangire, B. Sullivan, K. Fairchild, and V. Sampath, "Continuous vital sign analysis for predicting and preventing neonatal diseases in the twenty-first century: big data to the forefront," *Pediatric Research*, vol. 87, no. 2, pp. 210–220, 2020/01/01 2020, doi: 10.1038/s41390-019-0527-0.
- [28] A. E. W. Johnson *et al.*, "MIMIC-III, a freely accessible critical care database," *Scientific Data*, vol. 3, no. 1, p. 160035, 2016/05/24 2016, doi: 10.1038/sdata.2016.35.
- [29] C. Xiao, E. Choi, and J. Sun, "Opportunities and challenges in developing deep

- learning models using electronic health records data: A systematic review," *Journal of the American Medical Informatics Association*, vol. 25, no. 10, pp. 1419–1428, 2018/10// 2018, doi: 10.1093/jamia/ocy068.
- [30] M. Yang *et al.*, "Continuous prediction and clinical alarm management of late-onset sepsis in preterm infants using vital signs from a patient monitor," *Computer Methods and Programs in Biomedicine*, vol. 255, p. 108335, 2024/10/01/ 2024, doi: <https://doi.org/10.1016/j.cmpb.2024.108335>.
- [31] S. A. Collins *et al.*, "Relationship Between Nursing Documentation and Patients' Mortality," *American Journal of Critical Care*, vol. 22, no. 4, pp. 306–313, 2013, doi: 10.4037/ajcc2013426.
- [32] D. Agniel, I. S. Kohane, and G. M. Weber, "Biases in electronic health record data due to processes within the healthcare system: retrospective observational study," *BMJ*, vol. 361, p. k1479, 2018, doi: 10.1136/bmj.k1479.
- [33] B. J. Stoll *et al.*, "Trends in Care Practices, Morbidity, and Mortality of Extremely Preterm Neonates, 1993-2012," *JAMA*, vol. 314, no. 10, pp. 1039–1051, 2015, doi: 10.1001/jama.2015.10244.
- [34] Z. Peng *et al.*, "A Continuous Late-Onset Sepsis Prediction Algorithm for Preterm Infants Using Multi-Channel Physiological Signals From a Patient Monitor," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 1, pp. 550–561, 2023, doi: 10.1109/JBHI.2022.3216055.
- [35] J. C. Niestroy *et al.*, "Discovery of signatures of fatal neonatal illness in vital signs using highly comparative time-series analysis," *npj Digital Medicine*, vol. 5, no. 1, 2022/12// 2022, doi: 10.1038/s41746-021-00551-z.
- [36] B. A. Sullivan and K. D. Fairchild, "Vital signs as physiometers of neonatal sepsis," *Pediatric Research*, vol. 91, no. 2, pp. 273–282, 2022/1// 2022, doi: 10.1038/s41390-021-01709-x.
- [37] A. L. Schwab, B. Mayer, D. Bassler, H. D. Hummler, H. W. Fuchs, and M. B. Bryant, "Cerebral Oxygenation in Preterm Infants Developing Cerebral Lesions," (in English), *Frontiers in Pediatrics*, Original Research vol. Volume 10 - 2022, 2022–April–12 2022, doi: 10.3389/fped.2022.809248.
- [38] R. Lalitha *et al.*, "Multimodal Monitoring of Hemodynamics in Neonates With Extremely Low Gestational Age: A Randomized Clinical Trial," *JAMA Network Open*, vol. 8, no. 4, pp. e254101–e254101, 2025, doi: 10.1001/jamanetworkopen.2025.4101.
- [39] L. Letzkus *et al.*, "Heart rate patterns predicting cerebral palsy in preterm infants," *Pediatric Research*, vol. 97, no. 3, pp. 1040–1046, 2025/03/01 2025, doi: 10.1038/s41390-023-02853-2.
- [40] A. M. Zimmet *et al.*, "Vital sign metrics of VLBW infants in three NICUs: implications for predictive algorithms," *Pediatric Research*, vol. 90, no. 1, pp. 125–130, 2021/07/01 2021, doi: 10.1038/s41390-021-01428-3.
- [41] C. Dai, L. Buyu, X. Xin, and J. S. and Liu, "False Discovery Rate Control via Data Splitting," *Journal of the American Statistical Association*, vol. 118, no. 544, pp. 2503–2520, 2023/10/02 2023, doi: 10.1080/01621459.2022.2060113.
- [42] V. Berisha *et al.*, "Digital medicine and the curse of dimensionality," *npj Digital Medicine*, vol. 4, no. 1, p. 153, 2021/10/28 2021, doi: 10.1038/s41746-021-00521-5.
- [43] P. Ray, S. S. Reddy, and T. Banerjee, "Various dimension reduction techniques for high dimensional data analysis: a review," *Artificial Intelligence Review*, vol. 54, no. 5, pp. 3473–3515, 2021/06/01 2021, doi: 10.1007/s10462-020-09928-0.
- [44] W. Jia, M. Sun, J. Lian, and S. Hou, "Feature dimensionality reduction: a review," *Complex & Intelligent Systems*, vol. 8, no. 3, pp. 2663–2693, 2022/06/01 2022, doi:

- 10.1007/s40747-021-00637-x.
- [45] J. Song, Z. Li, G. Yao, S. Wei, L. Li, and H. Wu, "Framework for feature selection of predicting the diagnosis and prognosis of necrotizing enterocolitis," *PLOS ONE*, vol. 17, no. 8, p. e0273383, 2022, doi: 10.1371/journal.pone.0273383.
 - [46] A. Sheikhtaheri, M. R. Zarkesh, R. Moradi, and F. Kermani, "Prediction of neonatal deaths in NICUs: development and validation of machine learning models," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, p. 131, 2021/04/19 2021, doi: 10.1186/s12911-021-01497-8.
 - [47] N. Yalçın *et al.*, "Development and validation of a machine learning-based detection system to improve precision screening for medication errors in the neonatal intensive care unit," (in English), *Frontiers in Pharmacology*, Original Research vol. Volume 14 - 2023, 2023–April–14 2023, doi: 10.3389/fphar.2023.1151560.
 - [48] A. W. Manigault, S. J. Sheinkopf, H. F. Silverman, and B. M. Lester, "Newborn Cry Acoustics in the Assessment of Neonatal Opioid Withdrawal Syndrome Using Machine Learning," *JAMA Network Open*, vol. 5, no. 10, pp. e2238783–e2238783, 2022, doi: 10.1001/jamanetworkopen.2022.38783.
 - [49] L. Fregoso-Aparicio, J. Noguez, L. Montesinos, and J. A. García-García, "Machine learning and deep learning predictive models for type 2 diabetes: a systematic review," *Diabetology & Metabolic Syndrome*, vol. 13, no. 1, p. 148, 2021/12/20 2021, doi: 10.1186/s13098-021-00767-9.
 - [50] A. Thakkar and R. Lohiya, "Fusion of statistical importance for feature selection in Deep Neural Network-based Intrusion Detection System," *Information Fusion*, vol. 90, pp. 353–363, 2023/02/01/ 2023, doi: <https://doi.org/10.1016/j.inffus.2022.09.026>.
 - [51] P. Ghosh *et al.*, "Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques," *IEEE Access*, vol. 9, pp. 19304–19326, 2021, doi: 10.1109/ACCESS.2021.3053759.
 - [52] S. K. Thornton *et al.*, "UK and US risk factors for hearing loss in neonatal intensive care unit infants," *PLOS ONE*, vol. 19, no. 7, p. e0291847, 2024, doi: 10.1371/journal.pone.0291847.
 - [53] B. A. Sullivan *et al.*, "Comparing machine learning techniques for neonatal mortality prediction: insights from a modeling competition," *Pediatric Research*, 2024/12/16 2024, doi: 10.1038/s41390-024-03773-5.
 - [54] K. P. Shah, R.-A. O. deRegnier, W. A. Grobman, and A. C. Bennett, "Neonatal Mortality After Interhospital Transfer of Pregnant Women for Imminent Very Preterm Birth in Illinois," *JAMA Pediatrics*, vol. 174, no. 4, pp. 358–365, 2020, doi: 10.1001/jamapediatrics.2019.6055.
 - [55] C. J. Crilly, S. Haneuse, and J. S. Litt, "Predicting the outcomes of preterm neonates beyond the neonatal intensive care unit: What are we missing?," *Pediatric Research*, vol. 89, no. 3, pp. 426–445, 2021/02/01 2021, doi: 10.1038/s41390-020-0968-5.
 - [56] J. Lee, J. Cai, F. Li, and Z. A. Vesoulis, "Predicting mortality risk for preterm infants using random forest," *Scientific Reports*, vol. 11, no. 1, p. 7308, 2021/03/31 2021, doi: 10.1038/s41598-021-86748-4.
 - [57] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995, doi: <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
 - [58] B. Yoav and Y. Daniel, "The control of the false discovery rate in multiple testing under dependency," *The Annals of Statistics*, vol. 29, no. 4, pp. 1165–1188, 8/1 2001, doi: 10.1214/aos/1013699998.
 - [59] R. F. Barber and E. J. Candès, "Controlling the false discovery rate via knockoffs,"

- The Annals of Statistics*, vol. 43, no. 5, pp. 2055–2085, 31, 2015. [Online]. Available: <https://doi.org/10.1214/15-AOS1337>.
- [60] E. Candès, Y. Fan, L. Janson, and J. Lv, "Panning for Gold: 'Model-X' Knockoffs for High Dimensional Controlled Variable Selection," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 80, no. 3, pp. 551–577, 2018, doi: 10.1111/rssb.12265.
 - [61] W. Liu, Y. Ke, J. Liu, and R. Li, "Model-Free Feature Screening and FDR Control With Knockoff Features," *Journal of the American Statistical Association*, vol. 117, no. 537, pp. 428–443, 2022/01/02 2022, doi: 10.1080/01621459.2020.1783274.
 - [62] S. Bates, C. Emmanuel, J. Lucas, and W. and Wang, "Metropolized Knockoff Sampling," *Journal of the American Statistical Association*, vol. 116, no. 535, pp. 1413–1427, 2021/07/03 2021, doi: 10.1080/01621459.2020.1729163.
 - [63] M. Shao, X. Dong, and Y. and Zhang, "U-Statistic Reduction: Higher-Order Accurate Risk Control and Statistical-Computational Trade-Off," *Journal of the American Statistical Association*, pp. 1–14, doi: 10.1080/01621459.2024.2448029.
 - [64] J. Feng, J. Lee, Z. A. Vesoulis, and F. Li, "Predicting mortality risk for preterm infants using deep learning models with time-series vital sign data," *npj Digital Medicine*, vol. 4, no. 1, p. 108, 2021/07/14 2021, doi: 10.1038/s41746-021-00479-4.
 - [65] J. M. Gottman, "Time-series analysis," *A comprehensive introduction for social scientists*, 1981.
 - [66] B. D. Fulcher, M. A. Little, and N. S. Jones, "Highly comparative time-series analysis: the empirical structure of time series and their methods," *Journal of The Royal Society Interface*, vol. 10, no. 83, p. 20130048, 2013, doi: doi:10.1098/rsif.2013.0048.
 - [67] J. D. Hamilton, *Time series analysis*. Princeton university press, 2020.
 - [68] E. Parzen, "An approach to time series analysis," *The Annals of Mathematical Statistics*, vol. 32, no. 4, pp. 951–989, 1961.
 - [69] G. Marshall and L. Jonker, "An introduction to descriptive statistics: A review and practical guide," *Radiography*, vol. 16, no. 4, pp. e1–e7, 2010.
 - [70] R. J. Freund and W. J. Wilson, *Statistical methods*. Elsevier, 2003.
 - [71] P. Mishra, C. M. Pandey, U. Singh, A. Gupta, C. Sahu, and A. Keshri, "Descriptive statistics and normality tests for statistical data," *Annals of cardiac anaesthesia*, vol. 22, no. 1, pp. 67–72, 2019.
 - [72] W. Song, S. Y. Jung, H. Baek, C. W. Choi, Y. H. Jung, and S. Yoo, "A Predictive Model Based on Machine Learning for the Early Detection of Late-Onset Neonatal Sepsis: Development and Observational Study," (in English), *JMIR Med Inform*, Original Paper vol. 8, no. 7, p. e15965, 2020, doi: 10.2196/15965.
 - [73] J. Penttilä *et al.*, "Time domain, geometrical and frequency domain analysis of cardiac vagal outflow: effects of various respiratory patterns," *Clinical physiology*, vol. 21, no. 3, pp. 365–376, 2001.
 - [74] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, "Time Series Feature Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package)," *Neurocomputing*, vol. 307, pp. 72–77, 2018/09/13/ 2018, doi: <https://doi.org/10.1016/j.neucom.2018.03.067>.
 - [75] P. Bloomfield, *Fourier analysis of time series: an introduction*. John Wiley & Sons, 2004.
 - [76] P. Stoica and R. L. Moses, *Spectral analysis of signals*. Pearson Prentice Hall Upper Saddle River, NJ, 2005.
 - [77] M. B. Priestley, "The spectral analysis of time series," ed: Oxford University Press, 1988.

- [78] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bulletin of the American Meteorological society*, vol. 79, no. 1, pp. 61–78, 1998.
- [79] D. B. Percival and A. T. Walden, *Wavelet methods for time series analysis*. Cambridge university press, 2000.
- [80] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967, doi: 10.1109/TAU.1967.1161901.
- [81] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American journal of physiology-heart and circulatory physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.
- [82] C. Bandt and B. Pompe, "Permutation entropy: a natural complexity measure for time series," *Physical review letters*, vol. 88, no. 17, p. 174102, 2002.
- [83] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Physical review letters*, vol. 89, no. 6, p. 068102, 2002.
- [84] T. Burns and R. Rajan, "A Mathematical Approach to Correlating Objective Spectro-Temporal Features of Non-linguistic Sounds With Their Subjective Perceptions in Humans," (in English), *Frontiers in Neuroscience*, Original Research vol. Volume 13 - 2019, 2019–July–31 2019, doi: 10.3389/fnins.2019.00794.
- [85] A. Agresti and B. F. Agresti, "Statistical Analysis of Qualitative Variation," *Sociological Methodology*, vol. 9, pp. 204–237, 1978, doi: 10.2307/270810.
- [86] A. R. Wilcox, "Indices of Qualitative Variation," Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), 1967.
- [87] J. D. Teachman, "Analysis of Population Diversity: Measures of Qualitative Variation," *Sociological Methods & Research*, vol. 8, no. 3, pp. 341–362, 1980, doi: 10.1177/004912418000800305.
- [88] R. K. Peet, "The measurement of species diversity," *Annual review of ecology and systematics*, pp. 285–307, 1974.
- [89] D. Schleuter, M. Daufresne, F. Massol, and C. Argillier, "A user's guide to functional diversity indices," *Ecological Monographs*, vol. 80, no. 3, pp. 469–484, 2010, doi: <https://doi.org/10.1890/08-2225.1>.
- [90] G. D. Kader and M. Perry, "Variability for Categorical Variables," *Journal of Statistics Education*, vol. 15, no. 2, pp. null–null, 2007/07/01 2007, doi: 10.1080/10691898.2007.11889465.
- [91] J. P. Gibbs and D. L. Poston, Jr., "The Division of Labor: Conceptualization and Related Measures*," *Social Forces*, vol. 53, no. 3, pp. 468–476, 1975, doi: 10.1093/sf/53.3.468.
- [92] C. R. Harris *et al.*, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, 2020/09/01 2020, doi: 10.1038/s41586-020-2649-2.
- [93] P. Virtanen *et al.*, "SciPy 1.0: fundamental algorithms for scientific computing in Python," *Nature Methods*, vol. 17, no. 3, pp. 261–272, 2020/03/01 2020, doi: 10.1038/s41592-019-0686-2.
- [94] S. Seabold and J. Perktold, "Statsmodels: econometric and statistical modeling with python," *SciPy*, vol. 7, no. 1, pp. 92–96, 2010.
- [95] B. McFee *et al.*, "librosa: Audio and music signal analysis in python," *SciPy*, vol. 2015, pp. 18–24, 2015.
- [96] S. K. Lam, A. Pitrou, and S. Seibert, "Numba: a LLVM-based Python JIT compiler," presented at the Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC, Austin, Texas, 2015. [Online]. Available: <https://doi.org/10.1145/2833157.2833162>.

- [97] O. Manor and E. Segal, "Predicting Disease Risk Using Bootstrap Ranking and Classification Algorithms," *PLOS Computational Biology*, vol. 9, no. 8, p. e1003200, 2013, doi: 10.1371/journal.pcbi.1003200.
- [98] H.-Y. Lin *et al.*, "Cluster effect for SNP–SNP interaction pairs for predicting complex traits," *Scientific Reports*, vol. 14, no. 1, p. 18677, 2024/08/12 2024, doi: 10.1038/s41598-024-66311-7.
- [99] S. Sharma and S. Kumar, "Fast and accurate bootstrap confidence limits on genome-scale phylogenies using little bootstraps," *Nature Computational Science*, vol. 1, no. 9, pp. 573–577, 2021/09/01 2021, doi: 10.1038/s43588-021-00129-5.
- [100] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005, doi: <https://doi.org/10.1111/j.1467-9868.2005.00503.x>.
- [101] X. Xing, Z. Zhao, and J. S. Liu, "Controlling False Discovery Rate Using Gaussian Mirrors," *Journal of the American Statistical Association*, vol. 118, no. 541, pp. 222–241, 2023/01/02 2023, doi: 10.1080/01621459.2021.1923510.
- [102] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Physical Review E*, vol. 69, no. 6, p. 066138, 06/23/ 2004, doi: 10.1103/PhysRevE.69.066138.
- [103] F. J. Massey, "The Kolmogorov-Smirnov Test for Goodness of Fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951, doi: 10.2307/2280095.
- [104] R. Kerber, "ChiMerge: discretization of numeric attributes," presented at the Proceedings of the tenth national conference on Artificial intelligence, San Jose, California, 1992.
- [105] E. L. Fu, "Target Trial Emulation to Improve Causal Inference from Observational Data: What, Why, and How?," *Journal of the American Society of Nephrology*, vol. 34, no. 8, pp. 1305–1314, 2023, doi: 10.1681/asn.0000000000000152.
- [106] M. A. Hernán, W. Wang, and D. E. Leaf, "Target Trial Emulation: A Framework for Causal Inference From Observational Data," *JAMA*, vol. 328, no. 24, pp. 2446–2447, 2022, doi: 10.1001/jama.2022.21383.
- [107] R. A. Hubbard, C. A. Gatsonis, J. W. Hogan, D. J. Hunter, S.-L. T. Normand, and A. B. Troxel, "'Target Trial Emulation' for Observational Studies — Potential and Pitfalls," *New England Journal of Medicine*, vol. 391, no. 21, pp. 1975–1977, 2024, doi: doi:10.1056/NEJMp2407586.
- [108] B. Rasouli *et al.*, "Combining high quality data with rigorous methods: emulation of a target trial using electronic health records and a nested case-control design," *BMJ*, vol. 383, p. e072346, 2023, doi: 10.1136/bmj-2022-072346.
- [109] M. S. Mahmud, J. Z. Huang, S. Salloum, T. Z. Emara, and K. Sadatdiynov, "A survey of data partitioning and sampling methods to support big data analysis," *Big Data Mining and Analytics*, vol. 3, no. 2, pp. 85–101, 2020, doi: 10.26599/BDMA.2019.9020015.
- [110] J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, pp. 107–113, 2008, doi: 10.1145/1327452.1327492.
- [111] H. Wickham, "The Split-Apply-Combine Strategy for Data Analysis," *Journal of Statistical Software*, vol. 40, no. 1, pp. 1 – 29, 04/07 2011, doi: 10.18637/jss.v040.i01.
- [112] J. C. Anderson, J. Lehnardt, and N. Slater, *CouchDB: the definitive guide: time to relax.* " O'Reilly Media, Inc.", 2010.
- [113] J. Han, E. Haihong, G. Le, and J. Du, "Survey on NoSQL database," in *2011 6th international conference on pervasive computing and applications*, 2011: IEEE, pp.

- 363–366.
- [114] A. Moniruzzaman and S. A. Hossain, "Nosql database: New era of databases for big data analytics-classification, characteristics and comparison," *arXiv preprint arXiv:1307.0191*, 2013.
 - [115] S. Bradshaw, E. Brazil, and K. Chodorow, *MongoDB: the definitive guide: powerful and scalable data storage*. " O'Reilly Media, Inc.", 2019.
 - [116] V. Srinivasan *et al.*, "Aerospike: Architecture of a real-time operational dbms," *Proceedings of the VLDB Endowment*, vol. 9, no. 13, pp. 1389–1400, 2016.
 - [117] S. Sivasubramanian, "Amazon dynamoDB: a seamlessly scalable non-relational database service," in *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, 2012, pp. 729–730.
 - [118] J. R. Guay Paz, "Introduction to azure cosmos db," in *Microsoft Azure Cosmos DB Revealed: A Multi-Model Database Designed for the Cloud*: Springer, 2018, pp. 1–23.
 - [119] M. Zheludkov and T. Isachenko, *High Performance in-memory computing with Apache Ignite*. Lulu. com, 2017.
 - [120] J. Carpenter and E. Hewitt, *Cassandra: The Definitive Guide,(Revised)*. " O'Reilly Media, Inc.", 2022.
 - [121] B. Burns, J. Beda, K. Hightower, and L. Evenson, *Kubernetes: up and running: dive into the future of infrastructure*. " O'Reilly Media, Inc.", 2022.
 - [122] k. sigs. "kind." <https://kind.sigs.k8s.io/> (accessed).
 - [123] J. H. Friedman, "Multivariate Adaptive Regression Splines," *The Annals of Statistics*, vol. 19, no. 1, pp. 1–67, 67, 1991. [Online]. Available: <https://doi.org/10.1214/aos/1176347963>.
 - [124] M. Ali. "PyCaret: An open source, low-code machine learning library in Python." <https://www.pycaret.org> (accessed).
 - [125] H. Tin Kam, "Random decision forests," in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, 14–16 Aug. 1995 1995, vol. 1, pp. 278–282 vol.1, doi: 10.1109/ICDAR.1995.598994.
 - [126] H. F. Jerome, "Greedy function approximation: A gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 10/1 2001, doi: 10.1214/aos/1013203451.
 - [127] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997/08/01/ 1997, doi: <https://doi.org/10.1006/jcss.1997.1504>.
 - [128] P. E. van Beek, P. Andriessen, W. Onland, and E. Schuit, "Prognostic Models Predicting Mortality in Preterm Infants: Systematic Review and Meta-analysis," *Pediatrics*, vol. 147, no. 5, p. e2020020461, 2021, doi: 10.1542/peds.2020-020461.
 - [129] H. Al-Mandari *et al.*, "International survey on periextubation practices in extremely preterm infants," *Archives of Disease in Childhood - Fetal and Neonatal Edition*, vol. 100, no. 5, p. F428, 2015, doi: 10.1136/archdischild-2015-308549.
 - [130] D. Gupta *et al.*, "A predictive model for extubation readiness in extremely preterm infants," *Journal of Perinatology*, vol. 39, no. 12, pp. 1663–1669, 2019/12/01 2019, doi: 10.1038/s41372-019-0475-x.
 - [131] W. Song, Y. Hwa Jung, J. Cho, H. Baek, C. Won Choi, and S. Yoo, "Development and validation of a prediction model for evaluating extubation readiness in preterm infants," *Int J Med Inform*, vol. 178, p. 105192, Oct 2023, doi: 10.1016/j.ijmedinf.2023.105192.
 - [132] T. W. Rice, A. P. Wheeler, G. R. Bernard, D. L. Hayden, D. A. Schoenfeld, and L. B. Ware, "Comparison of the Spo2/Fio2 Ratio and the Pao2/Fio2 Ratio in Patients

- With Acute Lung Injury or ARDS," *CHEST*, vol. 132, no. 2, pp. 410–417, 2007, doi: 10.1378/chest.07-0617.
- [133] O. Roca *et al.*, "Predicting success of high-flow nasal cannula in pneumonia patients with hypoxemic respiratory failure: The utility of the ROX index," *Journal of Critical Care*, vol. 35, pp. 200–205, 2016/10/01/ 2016, doi: <https://doi.org/10.1016/j.jcrc.2016.05.022>.
 - [134] T. J. Pollard, A. E. W. Johnson, J. D. Raffa, and R. G. Mark, "tableone: An open source Python package for producing summary statistics for research papers," *JAMIA Open*, vol. 1, no. 1, pp. 26–31, 2018, doi: 10.1093/jamiaopen/ooy012.
 - [135] W. Song, Y. Hwa Jung, J. Cho, H. Baek, C. Won Choi, and S. Yoo, "Development and validation of a prediction model for evaluating extubation readiness in preterm infants," *International Journal of Medical Informatics*, vol. 178, p. 105192, 2023/10/01/ 2023, doi: <https://doi.org/10.1016/j.ijmedinf.2023.105192>.
 - [136] J. M. Di Fiore *et al.*, "The relationship between patterns of intermittent hypoxia and retinopathy of prematurity in preterm infants," *Pediatric Research*, vol. 72, no. 6, pp. 606–612, 2012/12/01 2012, doi: 10.1038/pr.2012.132.
 - [137] E. A. Jensen, S. B. DeMauro, M. Kornhauser, Z. H. Aghai, J. S. Greenspan, and K. C. Dysart, "Effects of Multiple Ventilation Courses and Duration of Mechanical Ventilation on Respiratory Outcomes in Extremely Low-Birth-Weight Infants," *JAMA Pediatrics*, vol. 169, no. 11, pp. 1011–1017, 2015, doi: 10.1001/jamapediatrics.2015.2401.
 - [138] B. Guy, M. E. Dye, L. Richards, S. O. Guthrie, and L. D. Hatch, "Association of time of day and extubation success in very low birthweight infants: a multicenter cohort study," *Journal of Perinatology*, vol. 41, no. 10, pp. 2532–2536, 2021/10/01 2021, doi: 10.1038/s41372-021-01168-6.
 - [139] W. Shalish, M. Keszler, P. G. Davis, and G. M. Sant'Anna, "Decision to extubate extremely preterm infants: art, science or gamble?," *Archives of Disease in Childhood - Fetal and Neonatal Edition*, vol. 107, no. 1, pp. 105–112, 2022, doi: 10.1136/archdischild-2020-321282.
 - [140] W. Shalish *et al.*, "The Impact of Time Interval between Extubation and Reintubation on Death or Bronchopulmonary Dysplasia in Extremely Preterm Infants," *The Journal of Pediatrics*, vol. 205, pp. 70–76.e2, 2019, doi: 10.1016/j.jpeds.2018.09.062.
 - [141] M. C. Walsh *et al.*, "Extremely Low Birthweight Neonates with Protracted Ventilation: Mortality and 18-Month Neurodevelopmental Outcomes," *The Journal of Pediatrics*, vol. 146, no. 6, pp. 798–804, 2005/06/01/ 2005, doi: <https://doi.org/10.1016/j.jpeds.2005.01.047>.
 - [142] P. A. Ahmann, A. Lazzara, F. D. Dykes, A. W. Brann Jr, and J. F. Schwartz, "Intraventricular hemorrhage in the high-risk preterm infant: Incidence and outcome," *Annals of Neurology*, vol. 7, no. 2, pp. 118–124, 1980, doi: <https://doi.org/10.1002/ana.410070205>.
 - [143] B. P. Murphy *et al.*, "Posthaemorrhagic ventricular dilatation in the premature infant: natural history and predictors of outcome," *Archives of Disease in Childhood - Fetal and Neonatal Edition*, vol. 87, no. 1, pp. F37–F41, 2002, doi: 10.1136/fn.87.1.F37.
 - [144] B. R. Vohr *et al.*, "Neurodevelopmental and Functional Outcomes of Extremely Low Birth Weight Infants in the National Institute of Child Health and Human Development Neonatal Research Network, 1993–1994," *Pediatrics*, vol. 105, no. 6, pp. 1216–1226, 2000, doi: 10.1542/peds.105.6.1216.
 - [145] J. D. Horbar *et al.*, "Trends in Mortality and Morbidity for Very Low Birth Weight Infants, 1991–1999," *Pediatrics*, vol. 110, no. 1, pp. 143–151, 2002, doi: 10.1542/peds.110.1.143.

- [146] Y. Futagi, Y. Toribe, K. Ogawa, and Y. Suzuki, "Neurodevelopmental Outcome in Children With Intraventricular Hemorrhage," *Pediatric Neurology*, vol. 34, no. 3, pp. 219–224, 2006, doi: 10.1016/j.pediatrneurol.2005.08.011.
- [147] L.-A. Papile, J. Burstein, R. Burstein, and H. Koffler, "Incidence and evolution of subependymal and intraventricular hemorrhage: A study of infants with birth weights less than 1,500 gm," *The Journal of Pediatrics*, vol. 92, no. 4, pp. 529–534, 1978, doi: 10.1016/S0022-3476(78)80282-0.
- [148] R. D. Sheth, "Trends in Incidence and Severity of Intraventricular Hemorrhage," *Journal of Child Neurology*, vol. 13, no. 6, pp. 261–264, 1998, doi: 10.1177/088307389801300604.
- [149] B. A. Darlow, A. E. Cust, and D. A. Donoghue, "Improved outcomes for very low birthweight infants: evidence from New Zealand national population based data," *Archives of Disease in Childhood - Fetal and Neonatal Edition*, vol. 88, no. 1, pp. F23–F28, 2003, doi: 10.1136/fn.88.1.F23.
- [150] J. M. Perlman, J. B. McMenamin, and J. J. Volpe, "Fluctuating Cerebral Blood-Flow Velocity in Respiratory-Distress Syndrome," *New England Journal of Medicine*, vol. 309, no. 4, pp. 204–209, 1983, doi: doi:10.1056/NEJM198307283090402.
- [151] J. J. Volpe, *Neurology of the newborn* (Major problems in clinical pediatrics, no. 22). Philadelphia: Saunders, 1981, pp. xv, 648 p.
- [152] M. Funato *et al.*, "Clinical events in association with timing of intraventricular hemorrhage in preterm infants," *The Journal of Pediatrics*, vol. 121, no. 4, pp. 614–619, 1992, doi: 10.1016/S0022-3476(05)81157-6.
- [153] I. L. Hand *et al.*, "Routine Neuroimaging of the Preterm Brain," *Pediatrics*, vol. 146, no. 5, 2020, doi: 10.1542/peds.2020-029082.
- [154] D. J. O. Starr R, Shah SD, *et al.*, *Periventricular and Intraventricular Hemorrhage*. StatPearls [Internet]: StatPearls Publishing, 2023 Aug 23.
- [155] K. K. Iyer *et al.*, "Early Detection of Preterm Intraventricular Hemorrhage From Clinical Electroencephalography," *Critical Care Medicine*, vol. 43, no. 10, pp. 2219–2227, 2015, doi: 10.1097/ccm.0000000000001190.
- [156] E. Von Elm, D. G. Altman, M. Egger, S. J. Pocock, P. C. Gøtzsche, and J. P. Vandenbroucke, "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies," *The lancet*, vol. 370, no. 9596, pp. 1453–1457, 2007.
- [157] W. A. Silverman and D. H. Andersen, "A CONTROLLED CLINICAL TRIAL OF EFFECTS OF WATER MIST ON OBSTRUCTIVE RESPIRATORY SIGNS, DEATH RATE AND NECROPSY FINDINGS AMONG PREMATURE INFANTS," *Pediatrics*, vol. 17, no. 1, pp. 1–10, 1956, doi: 10.1542/peds.17.1.1.
- [158] T. DelSole, "Optimally Persistent Patterns in Time-Varying Fields," (in English), *Journal of the Atmospheric Sciences*, vol. 58, no. 11, pp. 1341–1356, 01 Jun. 2001 2001, doi: [https://doi.org/10.1175/1520-0469\(2001\)058<1341:OPPITV>2.0.CO;2](https://doi.org/10.1175/1520-0469(2001)058<1341:OPPITV>2.0.CO;2).
- [159] D. J. Amaya *et al.*, "Bottom marine heatwaves along the continental shelves of North America," *Nature Communications*, vol. 14, no. 1, p. 1038, 2023/03/13 2023, doi: 10.1038/s41467-023-36567-0.
- [160] M. M. Qureshi *et al.*, "In vivo study of optical speckle decorrelation time across depths in the mouse brain," *Biomed. Opt. Express*, vol. 8, no. 11, pp. 4855–4864, 2017/11/01 2017, doi: 10.1364/BOE.8.004855.
- [161] F. Frascchetti and J. Giacalone, "EARLY-TIME VELOCITY AUTOCORRELATION FOR CHARGED PARTICLES DIFFUSION AND DRIFT IN STATIC MAGNETIC TURBULENCE," *The Astrophysical Journal*, vol. 755, no. 2, p. 114, 2012/08/01 2012, doi: 10.1088/0004-637X/755/2/114.
- [162] J. W. Bieber and W. H. Matthaeus, "Perpendicular Diffusion and Drift at

- Intermediate Cosmic-Ray Energies," *The Astrophysical Journal*, vol. 485, no. 2, p. 655, 1997/08/20 1997, doi: 10.1086/304464.
- [163] D. Wegierak *et al.*, "Decorrelation Time Mapping as an Analysis Tool for Nanobubble-Based Contrast Enhanced Ultrasound Imaging," *IEEE Transactions on Medical Imaging*, vol. 43, no. 6, pp. 2370–2380, 2024, doi: 10.1109/TMI.2024.3364076.
- [164] R. L. Hoiland, A. R. Bain, M. G. Rieger, D. M. Bailey, and P. N. Ainslie, "Hypoxemia, oxygen content, and the regulation of cerebral blood flow," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 310, no. 5, pp. R398–R413, 2016, doi: 10.1152/ajpregu.00270.2015.
- [165] J. M. J. R. Carr, R. L. Hoiland, I. A. Fernandes, W. G. Schrage, and P. N. Ainslie, "Recent insights into mechanisms of hypoxia-induced vasodilatation in the human brain," *The Journal of Physiology*, vol. 602, no. 21, pp. 5601–5618, 2024, doi: <https://doi.org/10.1113/JP284608>.
- [166] D. Quine and B. J. Stenson, "Arterial oxygen tension (Pao₂) values in infants <29 weeks of gestation at currently targeted saturations," *Archives of Disease in Childhood - Fetal and Neonatal Edition*, vol. 94, no. 1, pp. F51–F53, 2009, doi: 10.1136/adc.2007.135285.
- [167] M. J. Huizing, E. Villamor-Martínez, M. Vento, and E. Villamor, "Pulse oximeter saturation target limits for preterm infants: a survey among European neonatal intensive care units," *European Journal of Pediatrics*, vol. 176, no. 1, pp. 51–56, 2017/01/01 2017, doi: 10.1007/s00431-016-2804-9.
- [168] M. Kluckow and N. Evans, "Low superior vena cava flow and intraventricular haemorrhage in preterm infants," *Archives of Disease in Childhood - Fetal and Neonatal Edition*, vol. 82, no. 3, pp. F188–F194, 2000, doi: 10.1136/fn.82.3.F188.
- [169] A. G. Cimatti *et al.*, "Cerebral Oxygenation and Autoregulation in Very Preterm Infants Developing IVH During the Transitional Period: A Pilot Study," (in English), *Frontiers in Pediatrics*, Original Research vol. Volume 8 - 2020, 2020–July–15 2020, doi: 10.3389/fped.2020.00381.
- [170] N. J. Evans and L. N. Archer, "Postnatal circulatory adaptation in healthy term and preterm neonates," *Archives of Disease in Childhood*, vol. 65, no. 1 Spec No, pp. 24–26, 1990, doi: 10.1136/adc.65.1_Spec_No.24.
- [171] F. J. Walther, M. J. Benders, and J. O. Leighton, "Early changes in the neonatal circulatory transition," *The Journal of Pediatrics*, vol. 123, no. 4, pp. 625–632, 1993, doi: 10.1016/S0022-3476(05)80966-7.
- [172] S. Shankaran, C. R. Bauer, R. Bain, L. L. Wright, and J. Zachary, "Prenatal and Perinatal Risk and Protective Factors for Neonatal Intracranial Hemorrhage," *Archives of Pediatrics & Adolescent Medicine*, vol. 150, no. 5, pp. 491–497, 1996, doi: 10.1001/archpedi.1996.02170300045009.
- [173] A. Hansen and A. Leviton, "Labor and delivery characteristics and risks of cranial ultrasonographic abnormalities among very-low-birth-weight infants," *American Journal of Obstetrics & Gynecology*, vol. 181, no. 4, pp. 997–1006, 1999, doi: 10.1016/S0002-9378(99)70339-X.
- [174] A. Leviton, K. C. Kuban, L. Van Marter, M. Pagano, and E. N. Allred, "Antenatal Corticosteroids Appear to Reduce the Risk of Postnatal Germinal Matrix Hemorrhage in Intubated Low Birth Weight Newborns," *Pediatrics*, vol. 91, no. 6, pp. 1083–1088, 1993, doi: 10.1542/peds.91.6.1083.
- [175] S. E. Kolnik, K. Upadhyay, T. R. Wood, S. E. Juul, and G. C. Valentine, "Reducing Severe Intraventricular Hemorrhage in Preterm Infants With Improved Care Bundle Adherence," *Pediatrics*, vol. 152, no. 3, 2023, doi: 10.1542/peds.2021-056104.
- [176] K. C. K. Kuban *et al.*, "Neonatal Intracranial Hemorrhage and Phenobarbital,"

- Pediatrics*, vol. 77, no. 4, pp. 443–450, 1986, doi: 10.1542/peds.77.4.443.
- [177] B. J. S. al-Haddad *et al.*, "Effectiveness of a care bundle for primary prevention of intraventricular hemorrhage in high-risk neonates: a Bayesian analysis," *Journal of Perinatology*, vol. 43, no. 6, pp. 722–727, 2023/06/01 2023, doi: 10.1038/s41372-022-01545-9.
 - [178] Y.-H. Yang *et al.*, "Predicting early mortality and severe intraventricular hemorrhage in very-low birth weight preterm infants: a nationwide, multicenter study using machine learning," *Scientific Reports*, vol. 14, no. 1, p. 10833, 2024/05/12 2024, doi: 10.1038/s41598-024-61749-1.
 - [179] M. M. Farag, M. H. Gouda, A. M. A. Almohsen, and M. A. Khalifa, "Intraventricular hemorrhage prediction in premature neonates in the era of hemodynamics monitoring: a prospective cohort study," *European Journal of Pediatrics*, vol. 181, no. 12, pp. 4067–4077, 2022/12/01 2022, doi: 10.1007/s00431-022-04630-5.
 - [180] Y. Coskun, S. Isik, T. Bayram, K. Urgun, S. Sakarya, and I. Akman, "A clinical scoring system to predict the development of intraventricular hemorrhage (IVH) in premature infants," *Child's Nervous System*, vol. 34, no. 1, pp. 129–136, 2018/01/01 2018, doi: 10.1007/s00381-017-3610-z.
 - [181] J. B. Law *et al.*, "Intracranial Hemorrhage and 2-Year Neurodevelopmental Outcomes in Infants Born Extremely Preterm," *The Journal of Pediatrics*, vol. 238, pp. 124–134.e10, 2021, doi: 10.1016/j.jpeds.2021.06.071.
 - [182] Z. Nagy *et al.*, "Occurrence and Time of Onset of Intraventricular Hemorrhage in Preterm Neonates: A Systematic Review and Meta-Analysis of Individual Patient Data," *JAMA Pediatrics*, vol. 179, no. 2, pp. 145–154, 2025, doi: 10.1001/jamapediatrics.2024.5998.

Abstract in Korean

신생아 중환자실에 입원하는 저체중 미숙아는 높은 사망률과 이환율을 보이는 고위험 환자군에 속한다. 이러한 저체중 미숙아는 지속적인 모니터링과 높은 수준의 임상적 중재를 필요로 한다. 특히 미숙아의 생존율과 장기적인 예후를 개선하기 위해서는 조기 진단 및 예후 예측이 필수적이다. 이러한 요구를 만족하고 적합한 시점의 임상 결정과 중재를 지원하기 위해 최근 연속 활력징후 데이터를 기반의 임상 지표 식별 연구와 예측 모델 연구가 활발하게 진행되고 있다.

그러나 신생아 중환자실 환자군에 대한 머신러닝, 딥러닝 모델이 로지스틱 회귀 모델에 비해서 유의미한 우월성을 입증하지 못하고 있으며, 외부 검증 시에도 낮은 성능을 보이고 있다. 이러한 한계점은 다음과 같은 문제로 인해 발생하는 것으로 보고 되고 있다. 첫째, 생체신호 수집 및 처리 방식의 연구 및 기관별 차이로 인해 예측 모델의 일반화 가능성을 어렵게 하고 있다. 둘째, 미숙아의 재태 주수 및 임상 현장에서 발생하는 중재의 빈도가 기관 및 의료진 별로 상당한 이질성이 안정적인 지표 추출을 난해하게 만드는 것으로 알려지고 있다. 이와 더불어 연속적 활력징후를 활용하는 저체중 미숙아 연구는 높은 연산 부담과 시계열 분석 방법의 제한적 적용이라는 한계에 직면하여 어려움을 겪고 있다.

본 연구는 기존 저체중 미숙아 환자군의 특성으로 인한 분석의 어려움과 함께 분석 연산 자원 부담 및 시계열 분석 방법의 제한적인 적용이라는 문제, 그리고 기존의 연구 한계점 해결하기 위해 확장 가능한 연속 활력징후 분석 방법론을 제안하고자 한다. 이 방법론은

다양한 연구 영역의 시계열 분석 기법을 효율적으로 적용하여, 지속적으로 수집되는 대규모 활력징후 데이터로부터 임상적으로 연관성 있는 진단 및 예후 지표를 식별하고 새로운 생리적 요인을 심층적으로 탐색하도록 설계되었다.

특히, 이전에 발견되지 않은 새로운 연속 활력징후 기반 지표를 추출하기 위해 확장 가능한 특징 추출 접근 방식을 개발하여 방법론에 적용하였다. 전자의무기록 기반 데이터에서 식별하기 어려웠던 신생아중환자실 특유의 생리적 패턴을 반영하는 동적 특징을 도출하고자 시계열 분석 기법을 특징 추출에 통합했다. 또한, 최신 위양성발견을 제어 방법론과 임상 시험 에플레이션 방법을 분할 가능한 알고리즘으로 변환함으로써, 임상 지표 식별의 높은 확장성과 견고성을 동시에 향상시켰다. 이러한 알고리즘은 병렬 및 분산 컴퓨팅 기술 활용을 가능하게 하여, 고성능 컴퓨팅의 현재 추세에 발맞춰 대규모 다기관 임상 연구의 계산 효율성과 전반적인 확장성을 크게 높였다.

본 연구에서 제안한 방법론을 검증하기 위해 다음과 같은 연구를 수행하였다. 연구에서 제안한 위양성발견율이 기존의 방법론에 비해 더 높은 위양성발견을 제어와 연산 효율을 가지는지 확인하기 위해 시뮬레이션 분석을 통해 확인하였다. 기존의 저체중 미숙아의 주요 합병증인 패혈증과 사망 예측모델을 연구에서 제안하는 방법론을 기반으로 생성, 예측모델로 개발하였으며 외부 검증 데이터셋에서도 견고한 분류 성능을 보여주는 것을 확인하였다. 연속 활력징후 기반 예측 모델을 본 연구의 방법론을 기반으로 개발하여 기존의 의료진의 의사 결정에 기여를 할 수 있음을 확인하였다. 마지막으로 타 연구

영역의 시계열 분석 방법론을 활용하여 뇌실내출혈의 식별을 위한 새로운 지표를 본 방법론을 기반으로 식별함으로써 본 연구의 방법론이 새로운 임상 지표를 찾을 데 유용함을 확인할 수 있었다.

본 연구의 기여는 다음과 같다. (1) 연속 활력징후 데이터로부터 고해상도 임상 지표를 체계적으로 도출하여 저체중 미숙아의 위험 인자 식별 및 예측에 사용되는 특성 계산 방법론의 범위와 정밀도를 확장하였다. (2) 저체중 미숙아의 생리적 특성을 반영하는 시계열 분석 프레임워크를 개발하여 기존 저체중 미숙아 생체신호 분석 방법론의 한계를 완화하였다. (3) 본 연구에서 제안하는 방법론 기반으로 개발된 모델을 외부 검증함으로써 신생아 중환자실 내 예측모델의 신뢰성과 재현성을 향상하였다. (4) 새로운 생리적 지표에 대한 심층 분석을 통해 생리적 특성과 자율 신경계 기능 손상과 같은 중요한 임상 현상을 연결함으로써 모델의 해석 가능성과 임상적 유용성을 개선하고자 하였다. 본 연구는 기존 연구의 연산 및 분석적 한계를 완화하였으며, 연속적 활력 징후 연구의 실질적인 적용 가능성을 높여 연구 편의성 증진 및 임상적 문제 해결에 기여할 것으로 기대한다. 나아가 향후 저체중 미숙아의 예후 평가를 개선하고, 신생아 중환자실 환경에서 신뢰 가능하고 임상적으로 적용가능한 인공지능 모델 개발에 기여할 것으로 기대한다.

주제어: 미숙아, 연속 활력징후, FDR 제어, 시계열 분석, 고차원 데이터, 병렬 컴퓨팅, 표적 임상시험 에뮬레이션

학번: 2022-37947