Graduate School of Natural Sciences

Master Applied Data Science

Analyzing Temporal Patterns of ICU Alarms: A Time- Series Clustering Approach

Master Thesis

Author:

Pejman Sabournia

Supervisors:

Dr. Lex van Loon

Ruben Zoodsma

Examiner:

Prof. Dr. A.P.J.M. (Arno) Siebes

Utrecht, July 2024

Acknowledgements

I would like to express my gratitude to my supervisors, Lex van Loon and Ruben Zoodsma, for their support and encouragement throughout the thesis process. Additionally, I am grateful to Wilhelmina Kinderziekenhuis (WKZ) hospital for granting us access to their data.

Abstract

Objective

This study employs a machine learning approach, namely time-series clustering with Dynamic Time Warping (DTW), to empirically identify subgroups of intensive care unit (ICU) patients and examine the relationships between different alarm types. Utilizing unsupervised learning algorithm, the study aims to uncover insights into temporal alarm patterns in large datasets without the need for labeled data. The primary objectives are to identify distinct patterns of patient complications within the initial hours of ICU stay, regardless of admission diagnosis, and to understand how different types of alarms interact over time.

Design

The time-series clustering was performed using data from Wilhelmina Kinderziekenhuis (WKZ) hospital. The patient population included individuals from the pediatric and neonatology departments.

Conclusions

The study found occurrences of simultaneous cardiovascular and pulmonary physiological alarms, suggesting a correlation between them. Additionally, patterns of stable alarms, followed by surges, provide early warnings for patient functional decline, helping to optimize resource allocation.

Keywords: *machine learning, intensive care unit, time-series clustering, dynamic time warping*

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The intensive care unit (ICU) is a place where patients receive continuous and intensive physiological monitoring. The critical care team uses these data to apply interventions based on the patients' physiological status and then monitors the responses to these interventions, which inform subsequent treatment decisions (Costa & Kahn, 2016). The creation of an ICU represents organizational innovation by providing care for the sickest inpatients (Costa & Kahn, 2016). Physically grouping patients into a single location allows them to benefit from providers with expertise in caring for critically ill (Vranas et al., 2017).

In addition, electronic devices play a critical role in modern ICUs. Continuous monitoring of patients' vital parameters, as one of the most essential technical components of the intensive care unit (ICU), significantly improves patient safety by alerting staff through an alarm when a parameter deviates from the normal range (Poncette et al., 2021). However, the high frequency of alarms in ICUs has led to a phenomenon known as alarm fatigue among healthcare professionals, particularly nurses (Cho et al., 2016; Sowan et al., 2016). Alarm fatigue occurs when healthcare providers become desensitized to alarms due to their volume, leading to potential risks of missing critical alarms and compromising patient safety (Rayan et al., 2024).

Studies have shown that alarm management programs can impact healthcare providers' alarm fatigue (Dee et al., 2022). Hospitals can reduce the overall alarm burden and improve the response to critical alarms by implementing interventions such as changes in default alarm settings, providing alarm management training, and improving alarm notification systems (Sowan et al., 2016; Dee et al., 2022).

Furthermore, the type of patient monitored in the ICU can influence the alarm rate, with certain patient characteristics and specific medical conditions contributing to a higher alarm load (Sinno et al., 2022). Understanding patient characteristics, such as age and sex, as well as admission characteristics is critical for developing tailored alarm management strategies that address the unique needs of different patient populations in the ICU. By integrating machine learning algorithms with evidence-based alarm management strategies, healthcare providers can enhance patient safety, optimize alarm responses, and improve outcomes in critical care settings.

This study used a machine learning approach to better understand ICU alarm data. Unsupervised learning algorithms are particularly effective for dealing with ICU alarm data. These algorithms can identify patterns and anomalies in large datasets without the need for labeled data. For instance, Ghazanfari et al. (2019) employed an unsupervised feature-learning approach to reduce the false alarm rate in ICUs.

This study aims to provide insights into temporal alarm patterns using a time-series clustering algorithm. Therefore, the central research questions of this study are as follows:

- How can temporal analysis of ICU alarm data help identify distinct patterns of patient complications within the initial hours of ICU stay, regardless of admission diagnosis?

- How can the temporal analysis of ICU alarm data help understand the relationship between different types of alarms?

Chapter 2 discusses the methods and underlying concepts. Chapter 3 describes the data used and the pre-processing steps taken. Chapters 4 and 5 respectively provide a description and discussion of the main results.

Chapter 2

Methods

2.1 The methodological framework

This study used a modified version of the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, a popular methodological framework used for data mining. It consists of six key phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment (Spruit et al., 2021). This framework has been widely adopted in both academic and industrial settings because of its effectiveness in guiding researchers and data scientists through various stages of a project, ensuring a systematic and organized workflow (Ayele, 2020).

The first stage focuses on understanding the context in which project objectives and requirements are identified. Following this stage, the data understanding stage involves initial exploration to familiarize the team with the datasets. Subsequently, the data preparation stage focuses on integrating, cleaning, and transforming the data to make them suitable for modeling. Moving on to the modeling stage, this phase involves selecting and applying various modeling techniques to build and assess models for the data. In the evaluation stage, the developed models were evaluated to ensure that they effectively met the objectives. Finally, the deployment stage involves deploying the model and ensuring its proper functioning (Tunca, 2024). Figure 2.1 illustrates the modified version of the CRISP-DM used in this study.

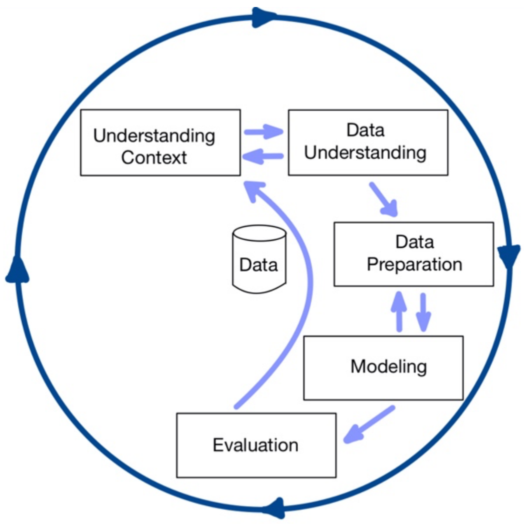

Figure 2.1: The adapted CRISP-DM framework used in this study

2.2 Core concepts

2.2.1 Time-series clustering

Unsupervised learning is a machine learning approach in which the model acquires patterns from the input data without any labeled answers. Unsupervised learning methods such as clustering are useful for tasks such as identifying anomalies, conducting exploratory data analysis, grouping clients, and uncovering correlations (Abdel-Besset et al., 2021). Clustering and dimensionality reduction are the primary algorithms used in unsupervised learning methods (Farhat et al., 2019).

Clustering involves grouping data points based on similarities in the data and is widely used in healthcare (Eckhardt et al., 2022). Similar to general clustering, time-series clustering groups similar objects. However, in time-series clustering, objects within the same cluster exhibit similar patterns over time, whereas those in different clusters display distinct temporal patterns (Zhu et al., 2023). This approach ensures that the clusters reflect both the inherent similarities and differences in how these patterns evolve over time, thereby providing deeper insight into the dynamics of the data.

In this study, time-series clustering was used for patient alarm data because it allows the comparison of patient alarm patterns over time. By analyzing the temporal patterns of alarms, clinicians can gain insights into the different progression patterns of patient conditions and explore the relationships between different alarm types.

2.2.2 Evaluation metrics

In this study, two evaluation metrics were used to determine the optimal number of clusters: within-cluster sum squares (WCSS) and the silhouette score. These metrics are essential for evaluating clustering quality and selecting the most suitable number of clusters.

The WCSS was calculated as the sum of the squared distances of the data points to their respective cluster centroids for each K value (Parkash et al., 2024). After calculating the WCSS for different numbers of clusters, an elbow plot is used to determine the optimal number of clusters. The elbow plot involves plotting the relationship between the number of clusters and

WCSS and selecting the point where there is the greatest rate of change in WCSS (Duggal et al., 2022).

The silhouette score is a widely used internal evaluation metric in cluster analysis, which assesses the quality of clustering by measuring the separation of clusters (Eliyanto & Surono, 2021). The silhouette score ranges from -1 to +1, where a score close to +1 indicates that the data points within a cluster are well matched and poorly matched to neighboring clusters, suggesting a good clustering result (Eliyanto & Surono, 2021).

In conclusion, determining the optimal number of clusters requires evaluating both WCSS and silhouette scores. Additionally, it is important to consider the interpretability of cluster numbers within the context of the study to ensure their practicality.

2.2.3 Dynamic time warping (DTW)

Dynamic Time Warping (DTW) is a classical dynamic programming algorithm that provides an optimal alignment between two time-series by non-linearly warping their time dimensions (Wang et al., 2014). Ratanamahatana and Keogh (2004) demonstrated its superiority over Euclidean distance in the classification and clustering of time-series data. DTW allows for the comparison of time-series data that may have variations in speed or timing, making it robust for matching similar patterns even when they are out of phase on the time axis (Tuzcu & Nas, 2005).

Patient alarms data patterns are often subject to various sources of noise such as sensor malfunctions, patient movements, or other external interferences. DTW's robustness allows it to focus on the core patterns of the time-series while minimizing the impact of these anomalies. Therefore, in the K-Means clustering algorithm, DTW is used as a metric to assess the similarity between various alarm patterns. This approach ensures that the clustering results are more accurate and reflect of the true underlying patterns in the data.

2.2.4 Min-Max normalization

Min-Max normalization is a common data pre-processing technique used in machine learning. This method involves scaling the values of a feature to a range between 0 and 1 by subtracting the minimum value and dividing it by the range of the data (Ampomah et al., 2021). This normalization technique is crucial as it ensures that all features contribute equally to the model training process, preventing any particular feature from dominating due to its large scale (Cao et al., 2016).

2.2.5 Kruskal-Wallis test

The Kruskal-Wallis test is a non-parametric statistical test that is used to compare two or more independent samples when the underlying population distribution is non-normal or unknown (Ostertagova et al., 2014). This test is particularly valuable in scenarios where parametric assumptions are not met, providing a robust alternative to parametric one-way ANOVA (Neve & Thas, 2015). By focusing on rank sums, the Kruskal-Wallis test allows researchers to determine whether samples originate from the same distribution, making it a versatile tool in various fields such as medicine, biometrics, and engineering (Sherwani et al., 2021).

In this study, the Kruskal-Wallis test was employed to determine the features that differed significantly between the clusters. This test is ideal for comparing features between clusters because the assumptions of normality and homogeneity of variances are not met. This analysis helps to understand the distinct characteristics of each cluster.

Chapter 3

Data

3.1 Ethical consideration

The alarm data used in this study were de-identified to ensure patient privacy and confidentiality. The de-identification of healthcare data is a crucial process that involves the removal of personally identified information from medical records to protect patient confidentiality and comply with privacy regulations (Trienes et al., 2020).

Additionally, access to the datasets was restricted to authorized personnel only, and the data analysis was conducted on a virtual machine approved by the Utrecht Medical Center. The study maintained patient privacy by adhering to ethical guidelines.

3.2 Data description

The alarm data for this study were collected from the Wilhelmina Kinderziekenhuis (WKZ) hospital, especially from two main departments: pediatric and neonatology. The pediatric department focuses on patients aged between 4 weeks and 18 years old, with most patients being between 4 weeks and 1 year old. In contrast, the neonatology department focuses on patients who are less than 4 weeks old.

The dataset spans approximately seven months, covering the period from October 1, 2023, to April 25, 2024. During this time, data were collected from 63 unique beds within the hospital, resulting in a dataset comprising approximately 3.9 million alarm entries. As shown in Figure 3.1, approximately 3 million entries were recorded from the neonatology department, while approximately 0.91 million entries recorded from the pediatric department.

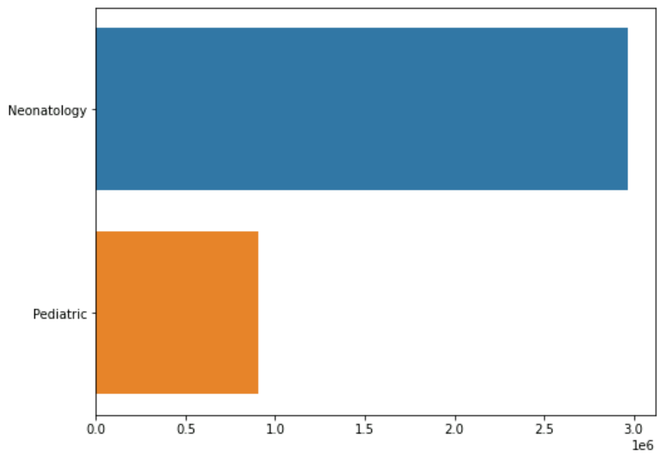

Figure 3.1: Distribution of alarm data by departments

The alarms were primarily generated using two types of equipment, patient monitors and ventilators, as shown in Figure 3.2. The alarms were categorized into two main groups: physiological and technical (Figure 3.3). Physiological alarms are related to the patient's vital signs and physiological state, whereas technical alarms are triggered by the status of medical equipment and patient artifacts such as movements or other activities that interfere with the sensors.

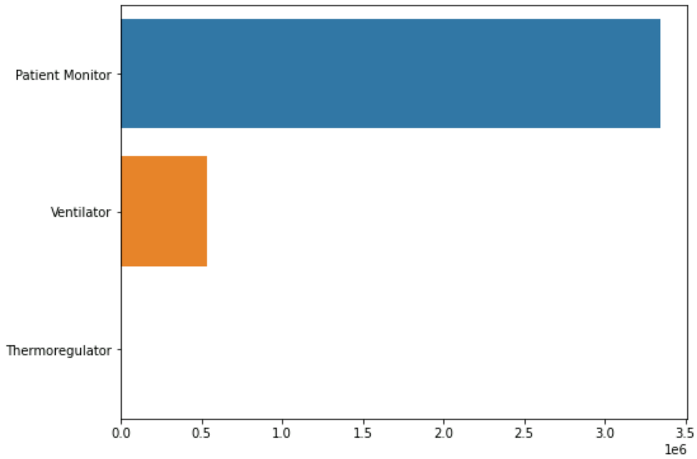

Figure 3.2: Distribution of alarms recorded from different devices

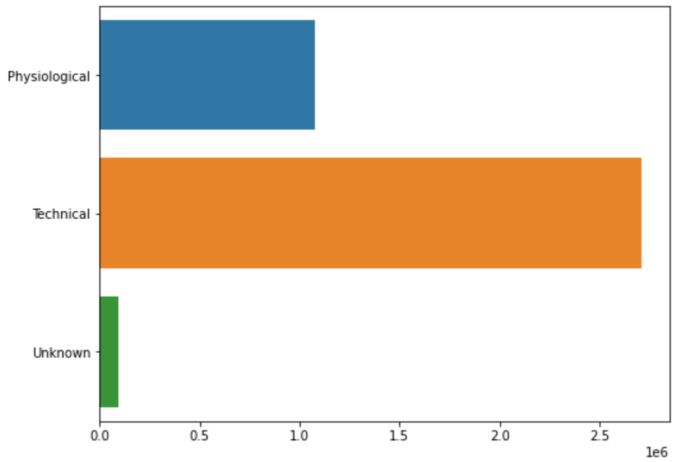

Figure 3.3: Distribution of alarms by type

3.3 Exploratory data analysis and pre-processing

Exploratory data analysis (EDA) and pre-processing of the ICU alarm data involved several steps to ensure that the dataset was ready for subsequent time-series clustering. Figure 3.4 shows the main steps taken to prepare the datasets. The process began with integration, where three distinct datasets were merged: the alarm dataset, which contained details about alarm start and end times, the devices generating the alarm, alarm duration in seconds, bed ID, and department; the patient dataset, which linked patient IDs to bed IDs; and alarm messages

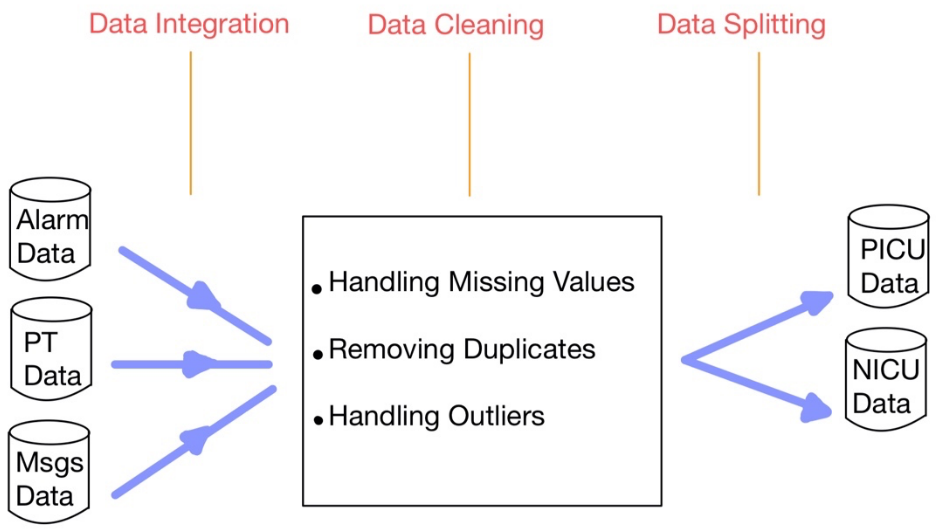

Figure 3.4: Data pre-processing flow

dataset that associated alarm messages with specific body organs. This integration provided a comprehensive dataset containing all necessary data for further analysis.

During the data cleaning stage, the dataset was examined to identify and address inconsistencies. Initially, missing values were addressed, and columns with more than 80% missing values, totaling 8 columns, were removed to maintain data integrity. Duplicate observations and columns were identified and eliminated. The exploratory data analysis revealed that there were abnormal alarm durations, with some negative values due to the device's time adjustments and others exceeding thousands of seconds, likely due to faulty devices. These anomalies were removed, and only observations within the 99th percentile of duration were retained. The cleaned dataset contained approximately 3.9 million observations and 26 columns. Figure 3.5 illustrates the distribution of alarm durations after the anomalies were removed.

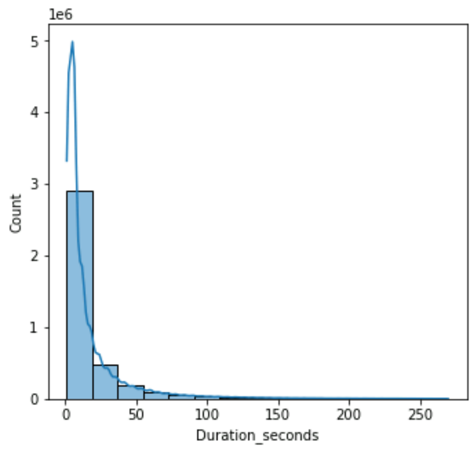

Figure 3.5: Distribution of alarm durations

Following the cleaning process, the dataset was split into two separate datasets based on the department: one for the pediatric intensive care unit (PICU) and another for the neonatology intensive care unit (NICU).

3.4 Additional pre-processing for modeling

In this study, additional pre-processing steps were performed to prepare the data for time-series clustering. Initially, for each dataset, PICU and NICU, a time window of 0.5 hours was created for each individual patient. Within each time window, the total number of alarms was counted for four specific features: pulmonary technical, pulmonary physiological, cardiovascular technical, and cardiovascular physiological.

Next, the data were filtered to include only the initial 12 hours of each patient's ICU stay, with the time windows organized in ascending order. This filtering focused on analyzing the critical initial period of ICU admission, when patient monitoring and intervention are most intensive. Maintaining a uniform observation window across all patients ensures comparability, which is important for obtaining robust clustering results.

The final pre-processing step involved extracting the alarm counts as arrays and applying the min-man normalization. The processed data were then used in the K-Means clustering algorithm, with dynamic time warping (DTW) as the metric for evaluating the similarities between different alarm patterns.

Chapter 4

Results

In this chapter, the results are presented in two parts: the first part focuses on the results from the pediatric ICU department and the second part shows the results from the neonatology ICU department.

4.1 PICU

4.1.1 Output of time-series clustering

As shown in Figure 4.1, the elbow point appears around the two clusters. This point represents a balance between reducing within-cluster sum squares (WCSS) and excessively partitioning the data into clusters with few observations. Although the elbow plot indicates that increasing the number of clusters beyond two continues to reduce the WCSS, the rate of improvement does not show a significant change. The silhouette scores for two and three clusters are 0.24 and 0.35, respectively. Although the silhouette score suggests that the three clusters may provide slightly better-defined clusters, two clusters were selected as the optimal number. This decision was made to avoid complicating clinical interpretations, as clusters containing only a few patients could result from increasing the number of clusters beyond two.

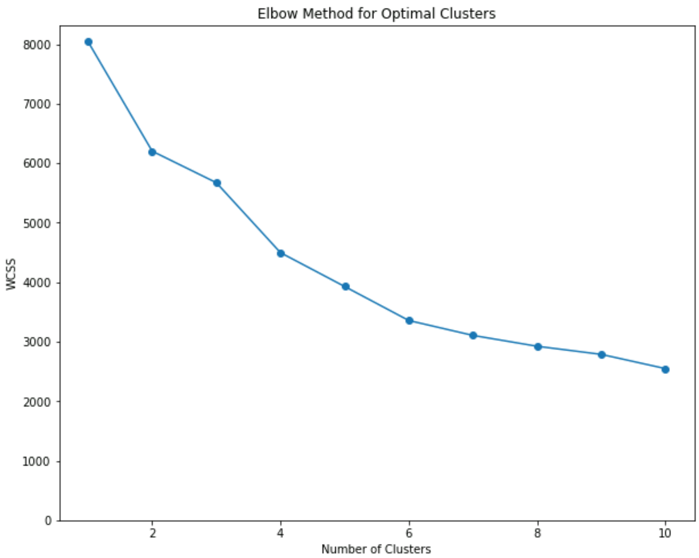

Figure 4.1: Elbow plot for determining the optimal number of clusters

Figure 4.2 displays the output of the time-series clustering process. The y-axis represents the alarm counts, and the x-axis shows time steps, where each time step corresponds to 0.5 hours, covering a total of 12 hours of ICU stay. Each line in the plots shows how the sum of the alarms in the predefined time steps varies for each patient across the selected features. Different colors represent the four features used in this analysis: pulmonary physiological, cardiovascular physiological, pulmonary technical, and cardiovascular technical alarms.

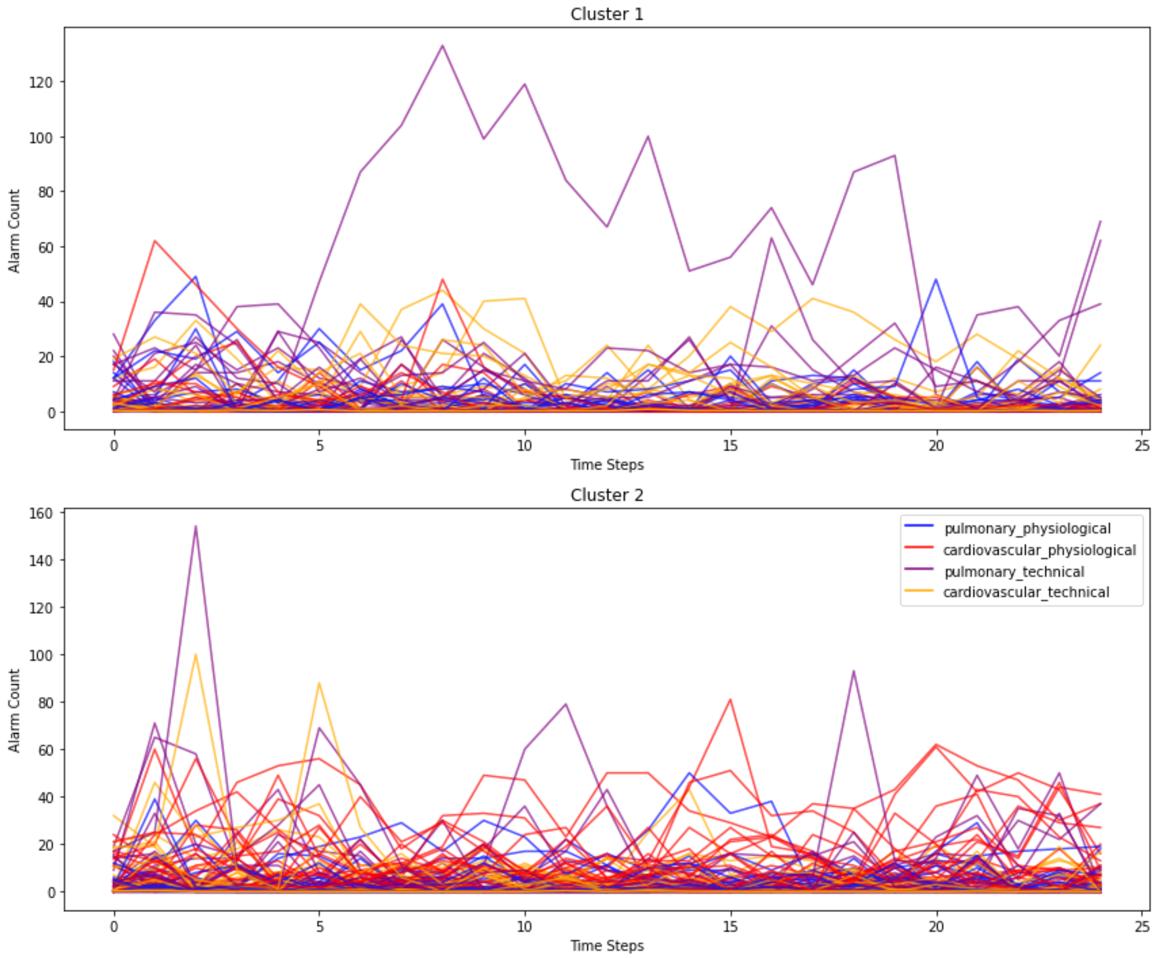

Figure 4.2: Alarm counts over time steps (0.5h) for each cluster - PICU

To gain deeper insights into the distinct patterns and behavior of each cluster, it is essential to examine the cluster centers (see Figure 4.3). These cluster centers show the normalized average of the alarm counts at each time step for each feature. Min-max normalization was used to scale the average alarm count between 0 and 1, allowing for a clearer comparison of patterns across different features.

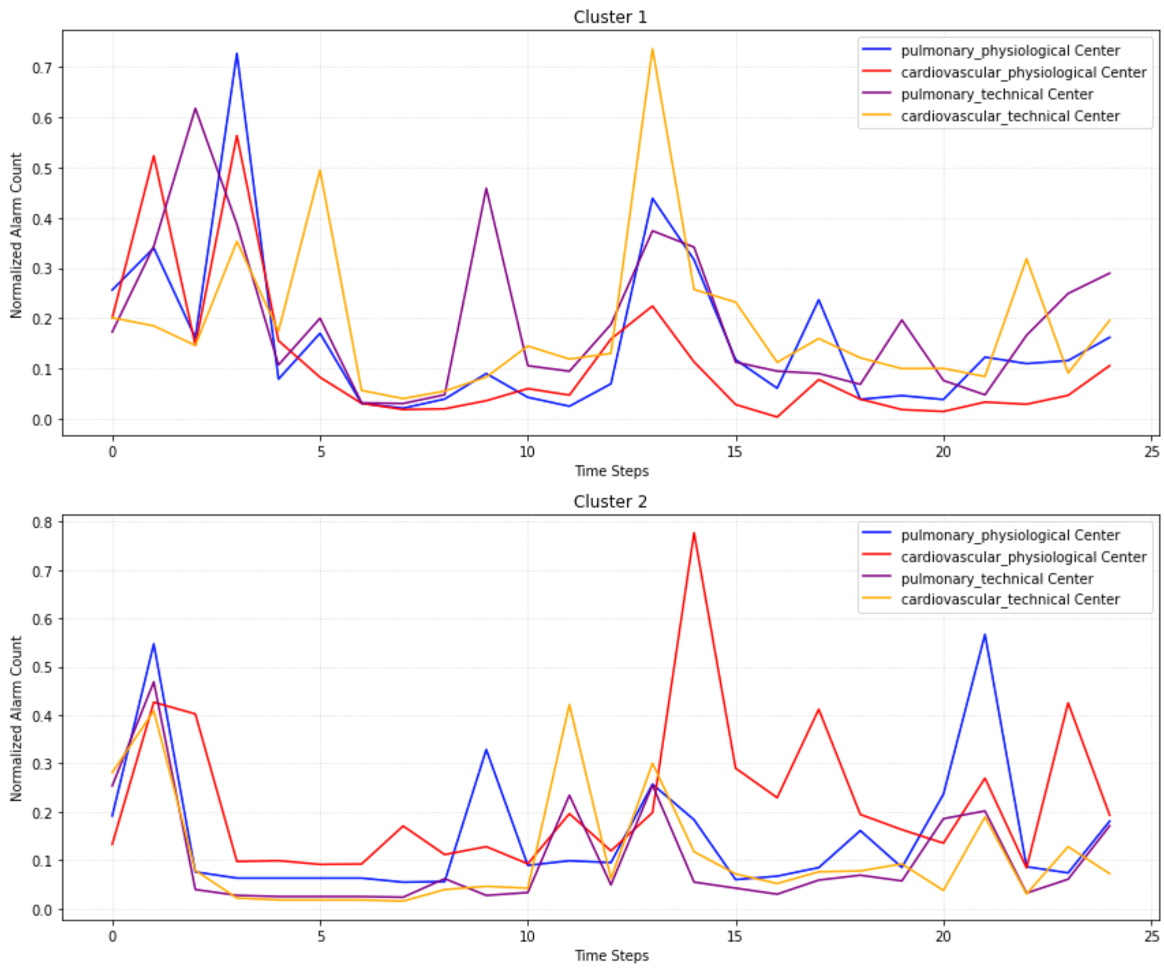

Figure 4.3: Normalized average alarm count over time steps (0.5 hour) for each cluster - PICU

In Cluster 1, which consists of 13 patients, the pulmonary physiological cluster center shows a noticeable peak at the 3^{rd} time step (2 hours of ICU stay), reaching a normalized alarm count of approximately 0.75. Following this peak, there is a sharp decline, with values remaining below 0.2, until another significant peak that occurs at the 13^{th} time step at 0.45. The cardiovascular physiological cluster center shows two distinct peaks at the 1^{st} and 3^{rd} time steps, reaching up to 0.55, followed by a sharp drop and almost remaining below 0.2 for the rest of the observation period. The pulmonary technical cluster center shows peaks at the 2^{nd}, 9^{th}, and 13^{th} time steps in decreasing order of magnitude, ending with an upward trend. The cardiovascular technical cluster center reaches a notable peak at the 13^{th} time step, approximately 0.75, followed by a sharp decline and several moderate peaks thereafter. By the end of the observation period, the cardiovascular physiological cluster center had the lowest value at 0.1, followed by the pulmonary physiological cluster center at approximately 0.15, the cardiovascular technical cluster center at 0.2, and the pulmonary technical cluster center at 0.3.

Additionally, there are specific time steps, such as the 3^{rd}, 5^{th}, and 13^{th}, where all features reach local peaks simultaneously.

In Cluster 2, which consists of 26 patients, the pulmonary physiological cluster center shows two noticeable peaks at the 1^{st} and 21^{st} time steps, reaching a normalized alarm count of approximately 0.55. There are some moderate fluctuations between these peaks, but overall, they remain stable. The cardiovascular physiological cluster center reaches its highest value at the 1^{st} time step, approximately 0.42, followed by a sharp drop to 0.1, and remains almost stable until it reaches its second highest peak at the 14^{th} time step, approximately 0.8. After this peak, it decreases significantly and exhibits some moderate fluctuations until the end of the observation period. At the first time step, the pulmonary technical cluster center shows a peak, reaching 0.45, followed by a significant drop, remaining stable until the 10^{th} time step. Afterward, it shows some moderate peaks, but it remains below 0.2 until the end of the observation period. The cardiovascular technical cluster center reaches notable peaks at the 1^{st} and 11^{th} time steps, approximately 0.42. Between these two peaks, it remains stable and stays below 0.1. Observations reveal some moderate peaks after the second peak, but the values still remain lower than the value of the second peak. By the end of the observation period, the pulmonary physiological and pulmonary technical cluster centers show an upward trend, whereas the cardiovascular physiological and cardiovascular technical centers show a downward trend. Similar to the first cluster, there are specific time steps in which all the features reach local peaks simultaneously.

In general, the cluster centers for all features show more fluctuations in Cluster 1 compared to Cluster 2, and in the second cluster, there are more stable periods. Additionally, for almost all features, the peak values are higher in the first cluster compared to the second cluster.

4.1.2 Clinical interpretation of cluster centers

In Cluster 1, the early peaks in pulmonary and cardiovascular physiological alarms, particularly at the 3^{rd} time step, may indicate an initial critical period shortly after ICU admission. This aligns with common clinical scenarios in which patients might be most unstable immediately after surgical interventions or acute events. The subsequent peaks, particularly those at the 13^{th} time step, where all four alarm types increase simultaneously, suggest recurring periods of instability that require close monitoring.

As time progresses, there is a noticeable increase in pulmonary physiological alarms, particularly after remaining low following the second peak. In general, the pulmonary physiological alarms are higher than the cardiovascular physiological alarms. Pulmonary technical alarms also have higher values compared to the cardiovascular physiological alarms. This suggests that patients in this cluster might experience increasing pulmonary complications as their stay in the ICU continues.

In Cluster 2, all types of alarms reach their first peak simultaneously shortly after ICU admission. After this initial phase, the alarm counts remain stable for an extended period, indicating these patients might stabilize more effectively compared to those in the first cluster. However, later in their ICU stay, there is a significant increase in cardiovascular physiological alarms. The cardiovascular physiological alarm counts are higher than the pulmonary physiological and technical alarms. Additionally, the cardiovascular physiological alarms fluctuate more than the pulmonary physiological alarms. This suggests that patients in this cluster might experience increasing cardiovascular complications as their stay in the ICU extends.

4.1.3 Statistical analysis

To further understand the differences between the identified clusters, a non-parametric statistical test was conducted. For the following analysis, the alarm counts for each patient within each cluster were individually summed for each feature. This summation provided an overview of the total alarms for each patient over the 12-hour ICU stay. The following statistical analysis was conducted to determine whether there were significant differences between alarm types across the clusters.

The Kruskal-Wallis test was used to examine the differences in alarm types for each feature across different clusters. An overview of the results can be found in Table 4.1.

Table 4.1: Kruskal-Wallis test results - PICU

Feature	**Statistic**	**p-value**
Pulmonary Physiological	3.58	0.058
Cardiovascular Physiological	7.76	0.0053
Pulmonary Technical	11.33	0.00076
Cardiovascular Technical	4.11	0.043

For pulmonary physiological alarms, the test statistic was 3.58 with a p-value of 0.058, indicating no significant difference between the clusters at the 5% significance level. In contrast, the remaining features: cardiovascular physiological, pulmonary technical and pulmonary technical showed significant differences across clusters, with test statistics of 7.76, 11.33, and 4.11, respectively, and p-values of 0.0053, 0.00076, and 0.043, repectively.

4.2 NICU

4.2.1 Output of time-series clustering

As shown in Figure 4.4, the elbow plot indicates that the rate of change was greatest at two. And the silhouette scores for two and three clusters are 0.35 and 0.28, respectively. Therefore, two clusters were selected as the optimal numbers.

Figure 4.5 displays the output of the time-series clustering process. The y-axis represents the alarm counts, and the x-axis shows time steps, where each time step corresponds to 0.5 hours, covering a total of 12 hours of ICU stay. Each individual line in the plots shows how the sum of the alarms in the predefined time steps varies for each patient across the selected features.

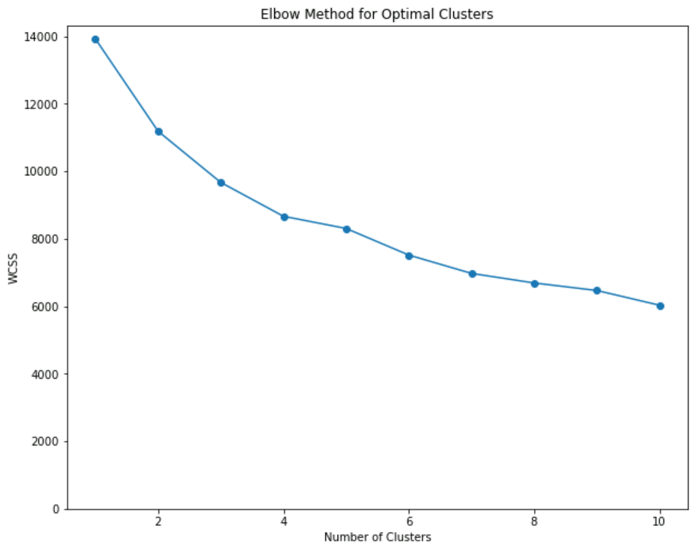

Figure 4.4: Elbow plot for determining the optimal number of clusters

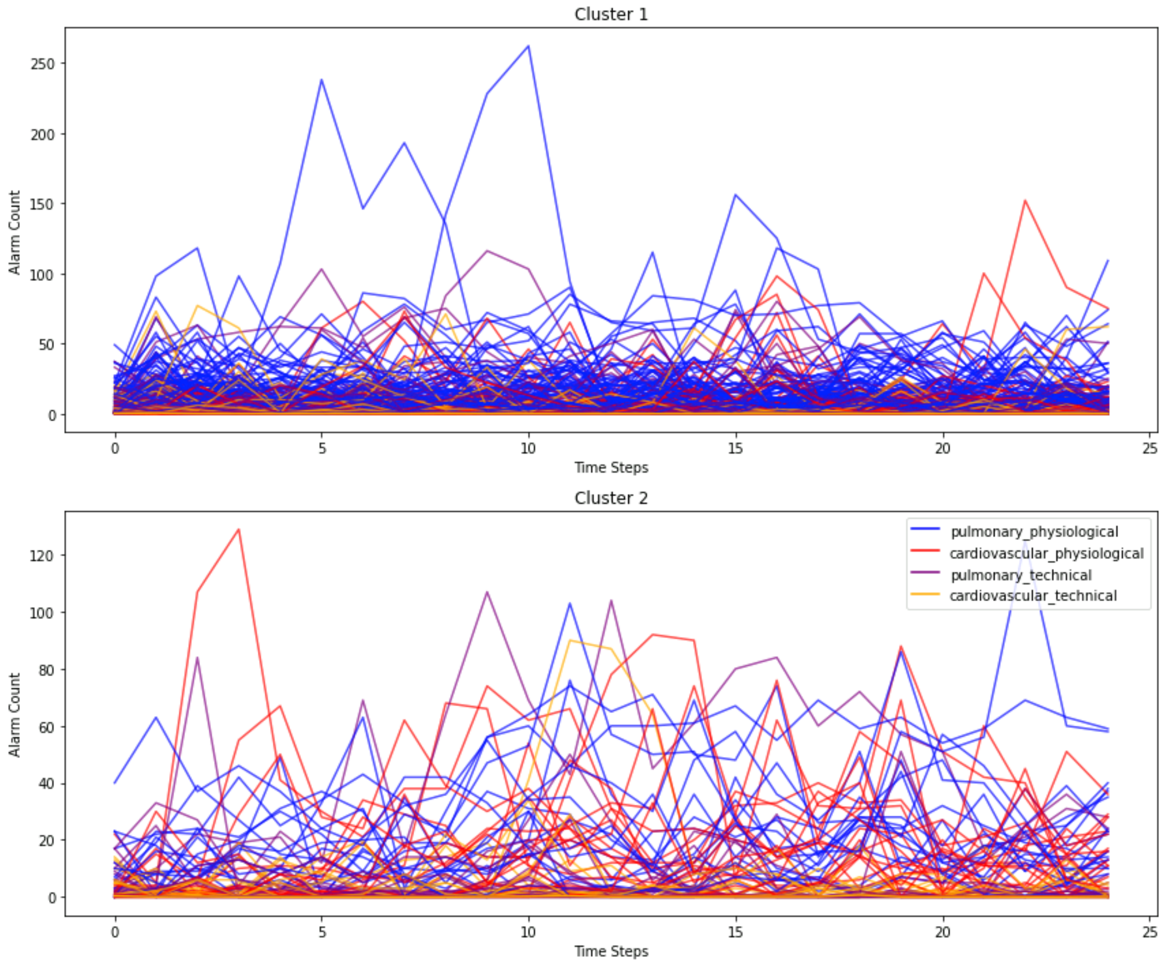

Figure 4.5: Alarm counts over time steps (0.5h) for each cluster - NICU

Similar to the pediatric clustering results, to gain deeper insights into the distinct patterns and behaviors of each cluster, it is essential to examine the cluster centers (see Figure 4.6). These cluster centers show the normalized average of alarm counts at each time step for each individual feature. Min-max normalization was used to scale the average of alarm counts between 0 and 1, allowing for a clearer comparison of patterns across different features.

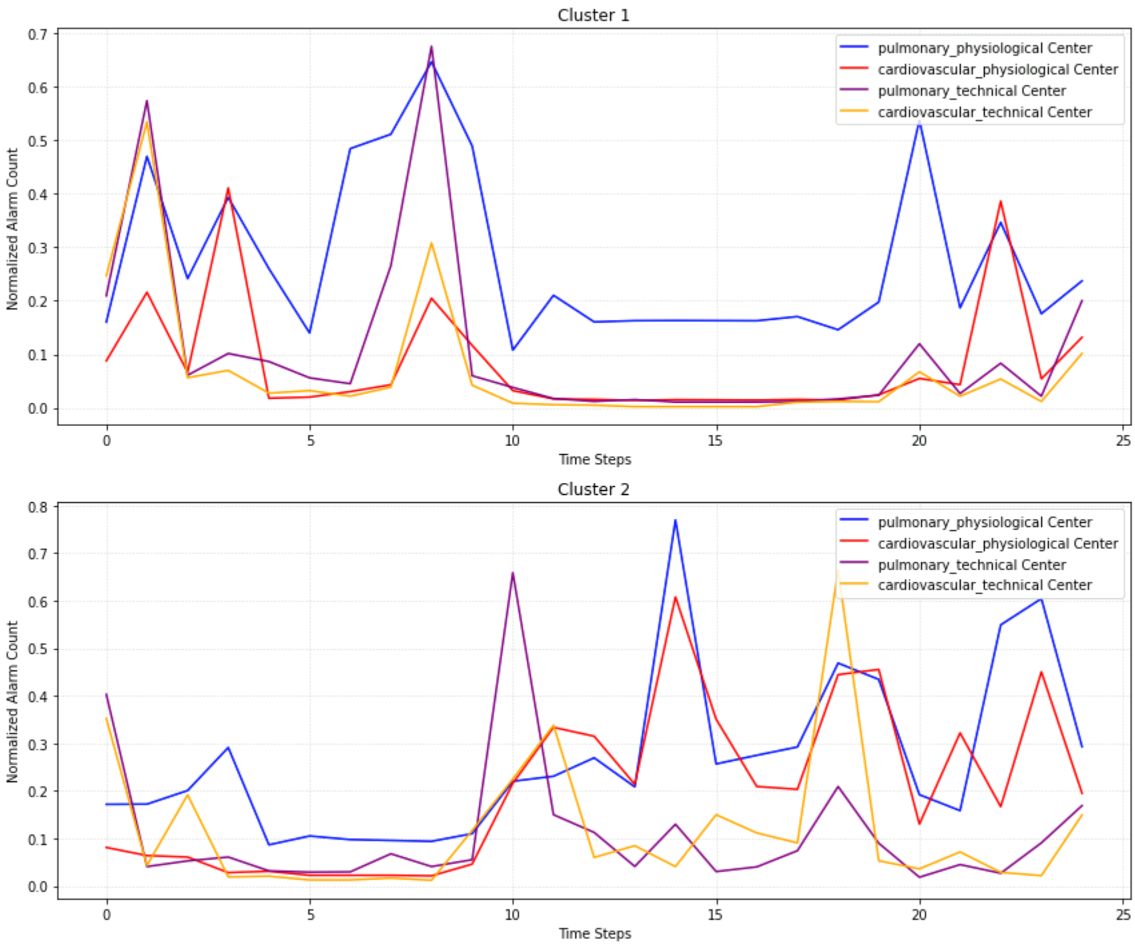

Figure 4.6: Normalized average alarm count over time steps (0.5 hour) for each cluster - NICU

In Cluster 1, which consists of 66 patients, the pulmonary physiological cluster center fluctuates at the early stages of admission, reaching a significant peak at approximately 0.65 at the 8th time step (4.5 hours of ICU stay), followed by a sharp drop below 0.2. It then remains stable for several hours until it begins to rise and fluctuate again, showing an upward trend at the end of the observation period. The other cluster centers—cardiovascular physiological, pulmonary technical, and cardiovascular technical—follow the same pattern as the pulmonary physiological. However, in general, their values are below pulmonary physiological alarm counts. Another noticeable observation is that all cluster centers reach their local maximum simultaneously.

In Cluster 2, which consists of 18 patients, the pulmonary physiological cluster center remains relatively stable and below 0.3 during the first half of the observation period. However, in the second half, it begins to rise and fluctuate. This pattern is similar for other alarm types, which also remain below 0.3 initially before increasing and fluctuating later. By the end of the observation period, both the cardiovascular and pulmonary physiological alarms show a downward trend.

4.2.2 Clinical interpretation of cluster centers

The analysis of Cluster 1 reveals significant fluctuations in the pulmonary physiological alarms during the first half of the observation period. This early instability is followed by a period of stability. Later, the alarm counts start to rise and fluctuate again, showing an upward trend toward the end, which could potentially indicate a worsening of the patient's condition. The simultaneous peaks in all cluster centers indicate periods during which patients' overall physiological condition became unstable. Additionally, the higher frequency of the pulmonary physiological alarms compared to other alarm types indicates the need for focused respiratory care in these patients.

Cluster 2 presents a different pattern; almost all of the alarm types initially are low and stable. This changes in the second half of the observation period, when all alarm types begin to rise and fluctuate. The downward trend of pulmonary and cardiovascular physiological alarms toward the end of the observation period may reflect recovery phases or effective interventions.

4.2.3 Statistical analysis

Similar to the PICU results, a non-parametric statistical test was conducted. For the following analysis, the alarm counts for each patient within each cluster were individually summed for each feature. This summation provided an overview of the total alarms for each patient over the 12-hour ICU stay. The following statistical analysis was conducted to determine whether there are significant differences between alarm types across the clusters.

The Kruskal-Wallis test was used to examine the differences in the alarm patterns for each feature across different clusters. An overview of the results can be found in Table 4.2.

Table 4.2: Kruskal-Wallis test results - NICU

Feature	Statistic	p-value
Pulmonary Physiological	0.56	0.45
Cardiovascular Physiological	24.5	7.39×10^{-7}
Pulmonary Technical	0.048	0.83
Cardiovascular Technical	3.58	0.059

For cardiovascular physiological alarms, the test statistic was 24.5 with a p-value of 7.39×10^{-7}, indicating a significant difference between clusters at the 5% significance level. In contrast, the remaining features: cardiovascular technical, pulmonary technical and physiological showed no significant differences across clusters, with test statistics of 3.58, 0.048, and 0.56, respectively, and p-values of 0.059, 0.83, and 0.45, respectively.

Chapter 5

Discussion, Conclusion, and Future Study

5.1 Discussion and conclusion

Given the evolving landscape of intensive care units (ICUs) and advanced monitoring systems in use, a deeper understanding of alarm systems is essential. This study aimed to uncover insights through a temporal analysis of ICU alarm data, particularly during the critical initial hours of a patient's stay. To achieve this, a time-series clustering algorithm combined with dynamic time warping was used as a metric to evaluate the similarities between alarm patterns across different patients.

One observation from the temporal pattern of the ICU alarm data was that, regardless of the department type, PICU or NICU, the temporal analysis identified distinct clusters of patient complications within the initial hours of ICU stay. One cluster included patients who experienced longer periods of complications early in their ICU stay, while another cluster included patients with shorter periods of complications. This distinction can help clinicians to group patients based on their immediate risk profiles. Recognizing which patients are likely to face prolonged complications regardless of their admission diagnosis, can prompt early and aggressive interventions. In contrast, patients with a shorter period of complications can be monitored with standard protocols, optimizing the use of ICU resources and ensuring focused attention where it is most needed.

Additionally, this study identified the simultaneous occurrence of cardiovascular and pulmonary physiological alarms, suggesting a correlation between these two types of alarms. Current monitoring systems may treat cardiovascular and pulmonary alarms independently, potentially leading to alarm fatigue due to numerous false positives. However, a spike in the heart rate accompanied by an abnormal respiratory rate is more likely to indicate a critical condition than changes in these metrics occurring independently. Therefore, by integrating data from both physiological alarms and recognizing their concurrent patterns, alarm systems can be refined to reduce the frequency of false alarms, ensuring that clinicians are alerted only when there is a genuine need for intervention.

In conclusion, this study underscores the importance of temporal analysis for understanding ICU alarm data. Recognizing patterns of simultaneous cardiovascular and pulmonary alarms, as well as periods of alarm surge, provides valuable early warnings for patient functional decline. These insights help clinicians optimize resource allocation, prioritize interventions, and tailor monitoring protocols to individual patient needs, and improve overall efficiency of ICU operations.

In addition, time-series clustering results can be used to forecast the deterioration of a patient's conditions. By tracking whether patients are downgraded or remain in the ICU after the initial hours of their stay could provide valuable data. These data can be used to determine whether the clustering results can be used to predict the deterioration of a patient's condition. Therefore, future studies should explore the predictive power of time-series clustering in forecasting patient outcomes and condition changes, thereby enhancing early intervention strategies and optimizing resource allocation in critical care settings.

5.2 Limitations

One limitation of this study was the absence of inclusion of patient attributes, such as age and gender. These factors have the potential to offer an additional understanding of the distinctions between clusters and can impact the interpretability of the clustering outcomes.

Another limitation arises from the existence of sample bias caused by the study's inclusion criteria. The investigation exclusively focused on patients who had a minimum duration of 12 hours in the intensive care unit. The use of this specific criterion may introduce bias by excluding patients with shorter stays in the ICU, who may have distinct alarm patterns and clinical progressions. As a result, the findings may not accurately represent the overall population of ICU patients, limiting the application of the results to a broader context.

Finally, the study was conducted in a specific intensive care unit context, and the observed patterns may not be generalizable to other ICUs with varied patient demographics and circumstances. Variances in patient demographics, hospital procedures, and medical apparatus may result in different alarm patterns.

5.3 Future studies

To gain a comprehensive understanding of the temporal analysis of ICU alarm data, future studies should consider the following factors. Employing random sampling methods is the initial stage to ensure that the sample accurately reflects the population and includes a variety of patient attributes. This strategy aims to mitigate sampling bias and enhance comprehension of ICU alarm dynamics by recording a diverse array of alarm patterns and clinical trajectories.

In addition, future studies should consider studying temporal alarm patterns for multiple ICUs with varied patient demographics and conditions. By combining data from many hospitals and locations, researchers may study the shared patterns and unique differences in temporal alarm patterns across diverse clinical contexts.

Bibliography

1. Costa, D. K., & Kahn, J. M. (2016). Organizing Critical Care for the 21st Century. JAMA, 315(8), 751–752. https://doi.org/10.1001/jama.2016.0974

2. Vranas, K. C., Jopling, J. K., Sweeney, T. E., Ramsey, M. C., Milstein, A. S., Slatore, C. G., Escobar, G. J., & Liu, V. X. (2017). Identifying Distinct Subgroups of ICU Patients: A Machine Learning Approach. Critical care medicine, 45(10), 1607–1615. https://doi.org/10.1097/CCM.0000000000002548

3. Poncette, A. S., Wunderlich, M. M., Spies, C., Heeren, P., Vorderwülbecke, G., Salgado, E., Kastrup, M., Feufel, M. A., & Balzer, F. (2021). Patient Monitoring Alarms in an Intensive Care Unit: Observational Study With Do-It-Yourself Instructions. Journal of medical Internet research, 23(5), e26494. https://doi.org/10.2196/26494

4. Cho, O., Kim, H., Lee, Y., & Cho, I. (2016). Clinical alarms in intensive care units: perceived obstacles of alarm management and alarm fatigue in nurses. Healthcare Informatics Research, 22(1), 46. https://e-hir.org/journal/view.php?id=10.4258/hir.2016.22.1.46

5. Sowan, A., Gomez, T., Tarriela, A., Reed, C., & Paper, B. (2016). Changes in default alarm settings and standard in-service are insufficient to improve alarm fatigue in an intensive care unit: a pilot project. Jmir Human Factors, 3(1), e1. https://humanfactors.jmir.org/2016/1/e1/

6. Rayan, A., Al-Ghabeesh, S.H., Fawaz, M., Behar, A., & Toumi, A. (2024). Experiences, barriers and expectations regarding current patient monitoring systems among icu nurses in a university hospital in lebanon: a qualitative study. Frontiers in Digital Health, 6. https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2024.1259409/full

7. Ghazanfari, B., Afghan, F., Najarian, K., & Mousavi, S. (2019). An Unsupervised Feature Learning Approach to Reduce False Alarm Rate in ICUs. https://www.researchgate.net/publication/332522104_An_Unsupervised_Feature_Learning_Approach_to_Reduce_False_Alarm_Rate_in_ICUs

8. Spruit, M., Kais, M., & Menger, V. (2021). Automated business goal extraction from e-mail repositories to bootstrap business understanding. *Future Internet, 13*(10), 243. https://www.mdpi.com/1999-5903/13/10/243

9. Ayele, W. (2020). Adapting CRISP-DM for Idea Mining. International Journal of Advanced Computer Science and Applications, 11(6). https://doi.org/10.14569/ijacsa.2020.0110603

10. Tunca, S. (2024). Forecasting the Enrolment of Bank Term Deposits: A case study approach with Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology. https://www.researchsquare.com/article/rs-3921578/v1

11. Xu, R. and Wunsch, D. (2005). Survey of Clustering Algorithms. Ieee Transactions on Neural Networks, 16(3), 645-678. https://ieeexplore.ieee.org/document/1427769

12. Parkash, B., Lie, T. T., Li, W., & Tito, S. R. (2024). End-to-end top-down load forecasting model for residential consumers. *Energies, 17*(11), 2550. https://www.mdpi.com/1996-1073/17/11/2550

13. Duggal, B., Duggal, M., Panch, A., Chourase, M., Gedam, P., Singh, P., & Subramanian, L. (2022). Using a national level cross-sectional study to develop a hospital preparedness index (hospi) for covid-19 management: a case study from india. Plos One, 17(7), e0269842. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0269842

14. Eliyanto, J. and Surono, S. (2021). Distance functions study in fuzzy c-means core and reduct clustering. Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika, 7(1), 118. https://doi.org/10.26555/jiteki.v7i1.20516

15. Wang, C., Alvarez, S., Ruiz, C., & Moonis, M. (2014). Semi-markov modeling-clustering of human sleep with efficient initialization and stopping.. https://doi.org/10.5220/0004824900610068

16. Ratanamahatana, C. and Keogh, E. (2004). Making time-series classification more accurate using learned constraints. https://epubs.siam.org/doi/10.1137/1.9781611972740.2

17. V. Tuzcu and S. Nas, "Dynamic time warping as a novel tool in pattern recognition of ECG changes in heart rhythm disturbances," *2005 IEEE International Conference on Systems, Man and Cybernetics*, Waikoloa, HI, USA, 2005, pp. 182-186 Vol. 1, doi: 10.1109/ICSMC.2005.1571142 https://ieeexplore.ieee.org/document/1571142

18. Rufener, C., Berezowski, J., Sousa, F., Abreu, Y., Asher, L., & Toscano, M. (2018). Finding hens in a haystack: consistency of movement patterns within and across individual laying hens maintained in large groups. Scientific Reports, 8(1). https://www.nature.com/articles/s41598-018-29962-x

19. Ampomah, E., Nyame, G., Qin, Z., Addo, P., Gyamfi, E., & Gyan, M. (2021). Stock market prediction with gaussian naïve bayes machine learning algorithm. Informatica, 45(2). https://doi.org/10.31449/inf.v45i2.3407

20. Cao, X., Stojković, I., & Obradović, Z. (2016). A robust data scaling algorithm to improve classification accuracies in biomedical data. BMC Bioinformatics, 17(1). https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-016-1236-x

21. Ostertagová, E., Ostertag, O., & Kováč, J. (2014). Methodology and application of the kruskal-wallis test. Applied Mechanics and Materials, 611, 115-120. https://doi.org/10.4028/www.scientific.net/amm.611.115
22. Neve, J. and Thas, O. (2015). A regression framework for rank tests based on the probabilistic index model. Journal of the American Statistical Association, 110(511), 1276-1283. https://www.tandfonline.com/doi/full/10.1080/01621459.2015.1016226

23. Sherwani, R., Shakeel, H., Awan, W., Faheem, M., & Aslam, M. (2021). Analysis of covid-19 data using neutrosophic kruskal wallis h test. BMC Medical Research Methodology, 21(1). https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-021-01410-x

24. Abdel-Basset, M., Moustafa, N., Hawash, H., & Ding, W. (2021). Unsupervised Deep Learning for Secure Internet of Things., 167-180. https://doi.org/10.1007/978-3-030-89025-4_6

25. Farhat, A., Shah, N., Wang, Z., & Raman, L. (2019). Machine learning: Brief overview for biomedical researchers. Journal of Translational Science, 6(3). https://www.oatext.com/machine-learning-brief-overview-for-biomedical-researchers.php#gsc.tab=0

26. Eckhardt, C., Madjarova, S., Williams, R., Ollivier, M., Karlsson, J., Pareek, A., & Nwachukwu, B. (2022). Unsupervised machine learning methods and emerging applications in healthcare. Knee Surgery Sports Traumatology Arthroscopy, 31(2), 376-381. https://link.springer.com/article/10.1007/s00167-022-07233-7

27. Zhu, C., Xu, Z., Lin, Z., Wang, L., Li, W., & Lu, S. (2023). Multivariate time series clustering based on graph convolutional network. Journal of Physics Conference Series, 2522(1), 012021. https://iopscience.iop.org/article/10.1088/1742-6596/2522/1/012021

28. Rai, P. & Singh, S. (2010). A survey of Clustering techniques, *International Journal of Computer Applications, 7*(12), 1-5. https://www.ijcaonline.org/volume7/number12/pxc3871808.pdf

29. Dee S.A., Tucciarone, J., Plotkin ,G., & Mallilo, C. Determining the Impact of an Alarm Management Program on Alarm Fatigue among ICU and Telemetry RNs: An Evidence Based Research Project. SAGE Open Nursing. 2022;8. doi:10.1177/23779608221098713 https://journals.sagepub.com/doi/10.1177/23779608221098713

30. Sinno, Z., Shay, D., Kruppa, J., Klopfenstein, S., Giesa, N., Flint, A., & Poncette, A. (2022). The influence of patient characteristics on the alarm rate in intensive care units: a retrospective cohort study. Scientific Reports, 12(1). https://www.nature.com/articles/s41598-022-26261-4

31. Trienes, J., Trieschnigg, D., Seifert, C., & Hiemstra, D. (2020). Comparing rule-based, feature-based and deep neural methods for de-identification of dutch medical records.. https://arxiv.org/abs/2001.05714

32. Vranas, K. C., Jopling, J. K., Sweeney, T. E., Ramsey, M. C., Milstein, A. S., Slatore, C. G., Escobar, G. J., & Liu, V. X. (2017). Identifying Distinct Subgroups of ICU Patients: A Machine Learning Approach. *Critical care medicine*, *45*(10), 1607–1615. https://doi.org/10.1097/CCM.0000000000002548

Appendix: Script Used

The following script was used in this project, and has been made available in the following link:
https://github.com/PejmanSa/ADS_thesis/blob/main/icu_alarm_analysis.ipynb

Name	Description
icu_alarm_analysis.ipynb	Pre-processing, EDA, and time-series clustering pipeline