Time Series Analysis of Hospital Admissions in the Brazilian Unified Health System (SIH/SUS): Trends in Morbidity, Mortality, and Hospital Costs in Santa Catarina

Victor R. L. M. de Almeida
Dep. de Informática e Estatística
Universidade Federal de Santa Catarina
Florianópolis, Brazil
victor.leite@posgrad.ufsc.br

Carina F. Dorneles
Dep. de Informática e Estatística
Universidade Federal de Santa Catarina
Florianópolis, Brazil
carina.dorneles@ufsc.br

Douglas D. J. de Macedo
Dep. de Informática e Estatística
Universidade Federal de Santa Catarina
Florianópolis, Brazil
douglas.macedo@ufsc.br

***Abstract*—This study presents a decade-long temporal analysis of hospital admissions recorded in the Brazilian Unified Health System (SIH/SUS) for the state of Santa Catarina from January 2015 to December 2024. Through an exploratory and data-driven approach, hospitalization indicators such as case fatality rate, average length of stay, total and mean hospital costs, and diagnostic composition were analyzed monthly over time. The results reveal clear impacts associated with the COVID-19 pandemic period (2020–2021), including a sharp increase in lethality (10.8%), a temporary rise in the average length of stay, and a peak in average hospitalization cost (R$ 2,768.48). Post-pandemic years show a gradual return to pre-2020 patterns, with stabilization of lethality around 3% and reduction in hospital costs. This longitudinal evaluation demonstrates the potential of SIH data for epidemiological surveillance, healthcare planning, and efficiency assessment, highlighting the role of open administrative datasets in supporting health intelligence in Brazil.**

***Index Terms*—SIH/SUS; Big Data in Healthcare; Hospital Admissions; Time Series Analysis; Healthcare Analytics; Hospital Mortality; Public Health Surveillance.**

I. Introduction

Administrative health information systems are essential tools for supporting evidence-based decision-making within public health systems worldwide. In Brazil, the *Hospital Information System of the Unified Health System* (SIH/SUS) stands out as one of the most comprehensive administrative hospital databases in existence, recording millions of inpatient admissions annually across all federated units [1], [2]. The SIH/SUS contains detailed information on demographic characteristics, diagnostic codes, procedures performed, hospitalization costs, and clinical outcomes, making it a strategic resource for monitoring healthcare utilization and evaluating the performance of the Brazilian health system.

Despite its breadth and longitudinal consistency, the SIH/SUS remains underexplored in analytical depth. Most published studies rely on cross-sectional or descriptive analyses, limiting the potential of the system to reveal long-term structural changes in morbidity, healthcare demand, and hospital mortality [3], [4]. However, the decade spanning 2015–2024 was marked by substantial transformations in the Brazilian healthcare landscape, particularly due to the COVID-19 pandemic. This period imposed unprecedented strain on hospital networks, significantly altering patterns of hospitalization, mortality, and financial expenditure [5], [6]. Understanding these temporal changes is crucial for assessing the resilience and vulnerabilities of the SUS.

The state of Santa Catarina (SC), located in southern Brazil, presents a relevant case for longitudinal analysis. With high socioeconomic indicators, an organized regionalized care network, and a diversified mix of public and philanthropic hospitals, Santa Catarina reflects both the strengths and challenges of the broader Brazilian health system [7]. Insights derived from temporal analyses in SC can therefore inform planning, epidemiological surveillance, and resource allocation at both state and national levels.

This study conducts a comprehensive temporal evaluation of SIH data from January 2015 to December 2024, covering nearly 120 months of hospital admissions. Through an integrated ETL (Extract–Transform–Load) pipeline and time-series exploratory analysis, we examine the evolution of key hospital indicators, including:

- hospital case fatality rate, an indicator of clinical severity and system pressure;
- average length of stay, reflecting case complexity and hospital operational efficiency;
- mean and total hospitalization costs, capturing financial burdens and reimbursement dynamics;
- monthly distribution of hospitalizations by ICD-10 diagnostic chapters;
- demographic profiles of admitted patients, including sex and age group distributions.

Beyond describing long-term trends, this study identifies

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

and contextualizes anomalous events, such as peaks in lethality and cost that coincide with critical epidemiological periods. The temporal characterization provided here establishes a foundation for future applications of machine learning and clustering techniques aimed at identifying latent patient profiles and hospital patterns within the SIH/SUS dataset [8], [9].

Overall, this work contributes to the growing field of healthcare analytics in Brazil by demonstrating how open, large-scale administrative databases such as the SIH/SUS can be leveraged to enhance public health intelligence, strengthen epidemiological monitoring, and support strategic planning in the context of a complex and evolving health system.

II. Related Work

The use of administrative hospital databases for epidemiological surveillance, health services research, and system performance evaluation has expanded considerably in recent decades [12]. In Brazil, the SIH/SUS has been the focus of several investigations due to its nationwide coverage and temporal consistency. Early studies emphasized descriptive analyses, data quality evaluation, and the characterization of disease-specific hospital utilization [3], [4]. These works highlighted both the strengths and limitations of SIH/SUS, noting challenges associated with coding practices, regional heterogeneity, and the lack of patient-level identifiers.

With advances in computational capacity and analytical methodologies, more recent research has begun to apply machine learning techniques to SIH/SUS data [13]. Examples include prediction models for hospital mortality, clustering for hospital efficiency assessment, and multivariate analyses of inpatient profiles [9]. Nonetheless, most of these studies examine short periods or restricted clinical domains, limiting their capacity to identify long-term systemic trends.

Internationally, hospital administrative datasets such as Medicare (USA), NHS (United Kingdom), and CIHI (Canada) have been widely analyzed using time-series methods, forecasting models, and unsupervised learning techniques [14]–[16]. These approaches have proven effective for monitoring healthcare demand, identifying seasonal fluctuations, and understanding the effects of large-scale events such as the COVID-19 pandemic [5], [6]. Such analyses demonstrate the value of longitudinal administrative data for predicting system stress, optimizing resource allocation, and evaluating policy interventions.

Time-series analyses have also been applied to study morbidity trends, age-related increases in hospital demand, and seasonal patterns in respiratory and circulatory diseases. However, few studies have systematically applied long-term temporal analysis to Brazilian administrative data [10]. The combined use of time-series visualization, exploratory analytics, and health system interpretation remains underdeveloped in the Brazilian literature.

In addition, international research on unsupervised learning—including clustering, dimensionality reduction, and anomaly detection—has shown promise in identifying latent utilization patterns in large healthcare datasets [11]. For instance, Van Der Laan et al. [8] demonstrated the utility of clustering for identifying patient subgroups in large-scale administrative data, supporting targeted clinical and managerial interventions.

Despite these contributions, there remains a significant gap in the literature regarding large-scale, decade-long analyses of SIH/SUS that integrate key hospital indicators such as lethality, length of stay, financial expenditure, diagnostic composition, and demographic structure into a unified analytical framework. Even fewer studies contextualize long-term temporal trends within specific state-level healthcare systems, such as Santa Catarina.

This work addresses these gaps by presenting a comprehensive ten-year temporal analysis (2015–2024) of hospitalization data from SIH/SUS using modern exploratory techniques. The findings lay the groundwork for the application of machine learning and clustering methods in future research.

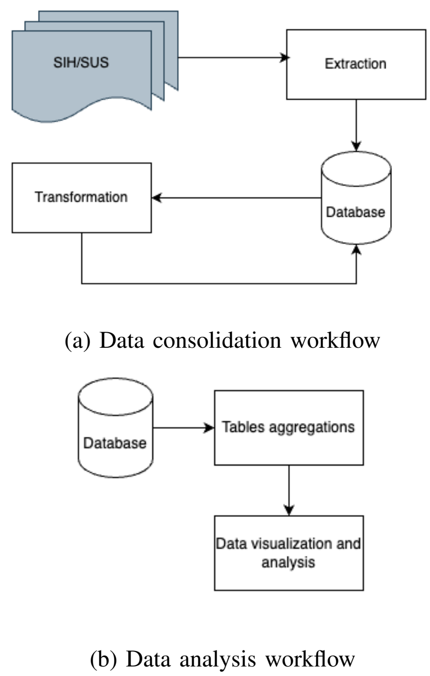

Fig. 1: Overview of the processing pipeline: (a) data consolidation, (b) data analysis.

III. Methodology

This study follows a data-driven analytical workflow designed to retrieve, process, and analyze ten years of administrative hospital data from the SIH/SUS system. The methodological pipeline consists of four main stages: (1) data acquisition, (2) ETL (Extract–Transform–Load) processing, (3) indicator computation, and (4) time-series exploratory analysis. All analyses were performed using Python in a Jupyter Notebook environment to ensure reproducibility.

A. Data Source and Acquisition

The dataset originates from the SIH/SUS, maintained by the Ministry of Health and distributed monthly by DATASUS [1], [2]. SIH/SUS records all hospital admissions reimbursed by the SUS. Data covering January 2015 to December 2024 (120 months) for the state of Santa Catarina were obtained using the PySUS Python library, which automates the download and conversion of monthly DBC files for Parquet files.

B. ETL Procedure

The ETL workflow adopted in this study is summarized in Figure 1. The process begins with the extraction of monthly SIH/SUS files, which are distributed in compressed `DBC` format by the Brazilian Ministry of Health. During the extraction phase (Figure 1a), the files are downloaded, decompressed, and converted into analyzable tabular structures. Once extracted, the records are loaded into an intermediate database that supports the subsequent transformation steps.

In the transformation stage, a series of data-cleaning and harmonization procedures are applied to ensure consistency across the ten-year historical series. These operations include handling missing or corrupted values, removing duplicated entries, standardizing ICD-10 diagnostic codes, normalizing monetary fields, and converting dates, numerical variables, and categorical descriptors into unified formats. Additional preprocessing steps include generating derived variables such as age groups, monthly temporal identifiers, and hospitalization severity indicators. After transformation, the standardized dataset is stored back into the database, where it becomes the foundation for the analytical phase.

The final stage of the workflow, illustrated in Figure 1b, focuses on the aggregation of indicators and computation of monthly metrics such as total hospitalizations, average length of stay, case fatality rate, and financial expenditures. These aggregated tables enable the subsequent visualization and analysis steps, which support time-series exploration and the identification of temporal morbidity trends.

The complete analysis pipeline is available online.[1]

C. Computed Indicators

The following monthly indicators were calculated for each transformation:

Hospital Case Fatality Rate:

$$\text{Fatality Rate} = \frac{\text{Number of deaths}}{\text{Total admissions}}$$

Average Length of Stay:

$$\text{ALOS} = \frac{\sum \text{Length of stay}}{\text{Total admissions}}$$

Mean Hospital Cost:

$$\text{Mean Cost} = \frac{\sum \text{Reimbursed value}}{\text{Total admissions}}$$

[1] https://github.com/vrodrigoleite/ufsc-ine-content-detection-paper

Total Hospital Expenditure:

$$\text{Total Cost} = \sum \text{Reimbursed value}$$

ICD-10 Chapter Distribution: Admissions were grouped by the first character of the ICD-10 code.

Sex and Age Group Distribution: Monthly counts were computed for sex and age categories.

D. Tools

Python 3.11, Polars, Pandas, NumPy, Matplotlib, Seaborn, and PySUS were used for all computations. Analyses were conducted in Jupyter Notebook.

E. Ethical Considerations

The SIH/SUS dataset contains only anonymized, public administrative data, and therefore complies with ethical guidelines for secondary data use.

IV. Results and Discussion

This section presents the exploratory analysis of SIH/SUS hospitalization data for the state of Santa Catarina from 2015 to 2024. The results are organized following the eight analytical dimensions proposed in this study, with figures referenced throughout the discussion.

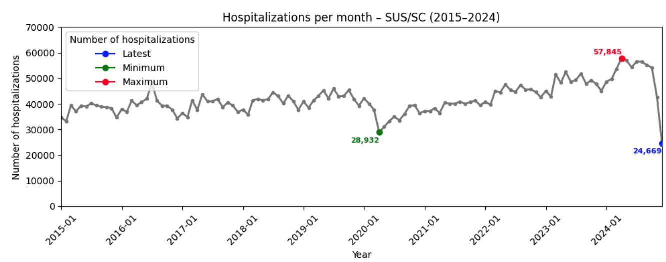

Fig. 2: Hospitalizations per month

A. Monthly Hospitalizations

The temporal evolution of total monthly hospitalizations is shown in Figure 2. The series exhibits a stable seasonal pattern from 2015 to early 2020, followed by a sharp decline coinciding with the onset of the COVID-19 pandemic, when elective procedures were suspended. A strong recovery is observed from late 2021 onward, surpassing pre-pandemic levels. The peak of the series occurs in early 2024, reaching 57,845 admissions, while the lowest point (28,932 admissions) aligns with early pandemic restrictions. The most recent value, 24,669 admissions, reflects seasonal behavior typical of end-of-year hospital dynamics.

B. Hospitalizations by Gender

Figure 3 presents monthly admissions disaggregated by gender. Female hospitalizations consistently exceed male hospitalizations throughout the period, suggesting a combination of higher utilization of reproductive health services, greater engagement with preventive care, and differences in health-seeking behaviors. Both groups show identical structural

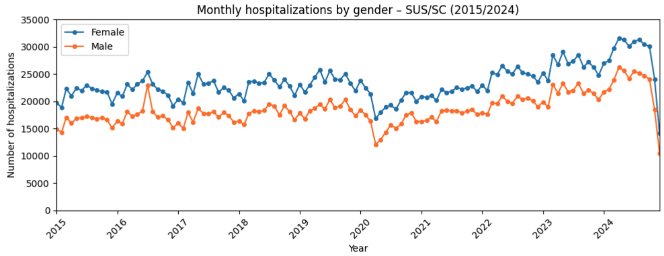

Fig. 3: Hospitalizations by gender

breaks during the pandemic, with identical drops in 2020, followed by a gradual return to normality. The widening gap observed after 2022 may reflect increased demand related to maternal health, chronic conditions among women, or recovery of suspended procedures.

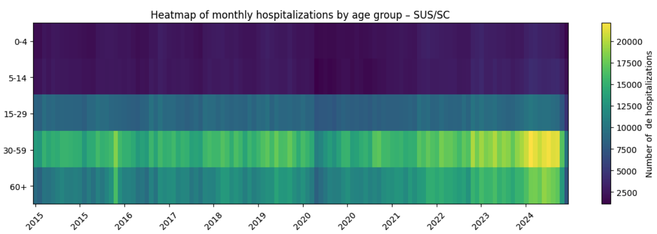

Fig. 4: Hospitalizations by gender

C. Hospitalizations by Age Group

The heatmap in Figure 4 illustrates hospitalization volume by age group. The 30–59 and 60+ categories dominate the overall demand, showing clear growth trends across the decade. This reflects demographic aging and the increasing burden of chronic diseases. The 60+ group, in particular, displays a strong upward trajectory from 2021 onward. Younger age groups (0–4 and 5–14) show stable and comparatively low hospitalization rates, consistent with global patterns of pediatric healthcare demand. The heatmap also captures the pandemic-related decline across all age segments in 2020, followed by heterogeneous recovery patterns.

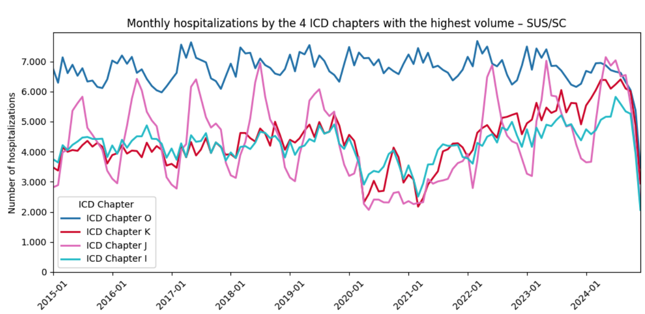

Fig. 5: Hospitalizations by most volume ICD-10 Chapter

D. Hospitalizations by ICD-10 Chapter

The distribution of hospitalizations across the four most frequent ICD-10 chapters (Figure 5) reveals heterogeneous temporal behaviors. Chapter O (pregnancy, childbirth, and puerperium) shows pronounced seasonality and remains the highest-volume category throughout the period. Chapters K (digestive system) and I (circulatory system) display steady growth aligned with population aging and increasing prevalence of chronic non-communicable diseases. Chapter J (respiratory diseases) presents the strongest fluctuations, with sharp declines during the pandemic due to social distancing effects that temporarily reduced respiratory infections. These dynamics reinforce the importance of segmenting hospital demand by clinical domain to understand structural drivers of utilization.

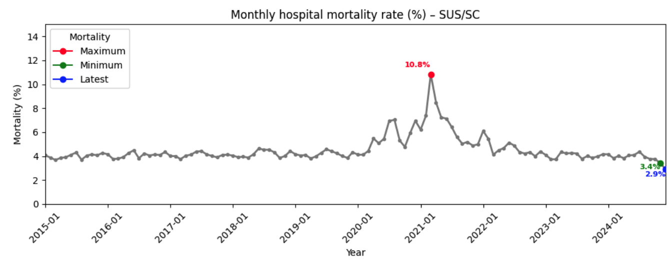

Fig. 6: Monthly Hospital Mortality Rate

E. Monthly Hospital Mortality Rate

Figure 6 shows the monthly hospital mortality rate. The series remains relatively stable (3%–5%) across most of the decade, with a dramatic peak in early 2021 (10.8%) corresponding to the second COVID-19 wave, particularly the Gamma variant surge in Brazil. After mid-2021, mortality gradually returns to pre-pandemic levels, reaching 2.9% in the most recent observation. The post-pandemic stabilization suggests normalization of care pathways and improved system resilience.

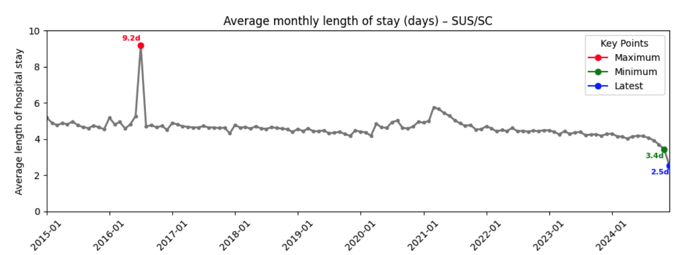

Fig. 7: Average monthly length of stay

F. Average Length of Stay

The average length of stay at Figure 7 demonstrates stable behavior across the observed period, with typical values around 4–5 days. An isolated spike to 9.2 days in 2016 is likely associated with data irregularities or a localized anomaly in hospitalization profiles for that month. During the pandemic, a modest increase is observed due to prolonged stays for severe COVID-19 cases. By 2024, the metric reaches its lowest values (2.5–3.4 days), which may indicate greater efficiency in case management, expanded use of outpatient pathways, or a shift in case mix.

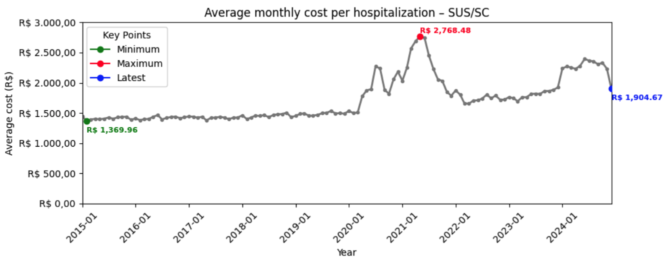

Fig. 8: Average monthly cost

G. Average Monthly Cost per Hospitalization

The mean cost per hospitalization (Figure 8) remains stable between 2015 and 2019, with values near R$ 1,400. During the pandemic, substantial volatility emerges, peaking at R$ 2,768 in early 2021 due to high clinical complexity, greater resource use, and elevated prices of medical supplies and intensive care. Post-2022, costs stabilize again, though at a higher baseline than the pre-pandemic period, reflecting structural increases in healthcare expenditures.

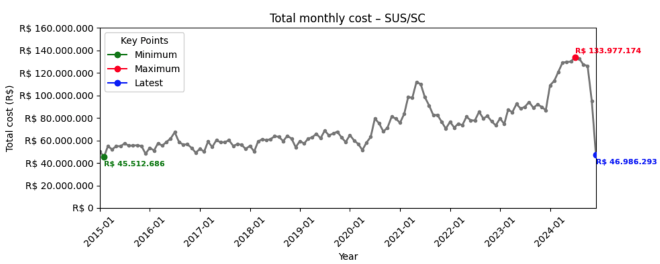

Fig. 9: Total monthly cost

H. Total Monthly Hospital Expenditure

The total monthly cost of hospitalizations (Figure 9) combines both utilization volume and unit cost. The trend shows modest growth until 2020, followed by large oscillations driven by pandemic pressures. The maximum value (R$ 133,977,174) occurs in early 2024, coinciding with high hospitalization volume. The most recent value (R$ 46,986,293) reflects the December seasonal decline in demand. Over the decade, the general trend indicates increasing financial pressure on the state healthcare system, partially driven by demographic aging and rising treatment complexity.

V. Conclusion

This study presented a comprehensive ten-year analysis of hospitalizations recorded in the SIH/SUS system for the state of Santa Catarina, covering the period from 2015 to 2024. By integrating a structured ETL workflow, robust data preprocessing, and multi-dimensional exploratory analytics, the work revealed long-term patterns in hospital utilization, demographic dynamics, clinical demand, and financial pressures within the state's public health system.

The results highlight several important insights. First, hospitalization volumes exhibit strong seasonal behavior and clear disruptions associated with the COVID-19 pandemic, followed by a rapid recovery that eventually surpassed pre-pandemic levels. Gender- and age-specific analyses showed persistent demand concentration among women and older adults, emphasizing the need for age-sensitive and gender-sensitive health policies. Clinical demand patterns based on ICD-10 chapters revealed structural drivers of hospitalizations, particularly those related to reproductive health, chronic diseases, and respiratory conditions.

Financial indicators demonstrated that both the average cost per hospitalization and total monthly expenditures increased substantially during and after the pandemic, reflecting rising clinical complexity and systemic cost inflation. Meanwhile, hospital mortality and average length of stay returned to stable pre-pandemic patterns, suggesting normalization of care pathways and improved post-pandemic resilience.

Overall, this study demonstrates the value of long-term administrative healthcare data for monitoring system behavior, identifying pressure points, and supporting evidence-based decision making. The dataset and methodological pipeline developed here provide a foundation for future research involving predictive modeling, clustering techniques, and causal inference to better understand the determinants of hospital utilization and optimize resource allocation within the SUS.

Metadata

<dc:title>Temporal Analysis of Hospital Admissions
<dc:creator>Almeida, Victor R. L. M. de</dc:creato
<dc:subject>SIH/SUS; Big Data in Healthcare; Hospi
<dc:description>This project presents a ten-year l
<dc:publisher>Universidade Federal de Santa Catari
<dc:date>2025-12</dc:date>
<dc:type>Text; Research article</dc:type>
<dc:format>application/pdf</dc:format>
<dc:identifier>https://github.com/vrodrigoleite/uf
<dc:source>Hospital Information System of the Braz
<dc:language>en</dc:language>
<dc:relation>Analysis code and notebooks</dc:relat

References

[1] BRASIL. Ministério da Saúde. DATASUS. *Sistema de Informações Hospitalares do SUS (SIH/SUS): Documentação e Bases de Dados*. Ministério da Saúde, Brasília, 2024.

[2] BRASIL. Ministério da Saúde. *Manual Técnico do Sistema de Informações Hospitalares – SIH/SUS*. Secretaria de Atenção à Saúde, Brasília, 2019.

[3] MACHADO, J. P.; MARTINS, M.; LEITE, I. C. "Quality of hospital administrative data in Brazil: a systematic review." *Cadernos de Saúde Pública*, vol. 34, no. 1, 2018.

[4] PESTANA, M. H.; ROCHA, M. S. "Hospital morbidity in Brazil: advances and challenges using SIH/SUS data." *Revista de Saúde Pública*, vol. 53, 2019.

[5] BASTOS, L. S.; NOBRE, F. F.; PEREIRA, R. H.; et al. “COVID-19 hospitalizations in Brazil: regional inequalities and system overload.” *The Lancet Regional Health – Americas*, vol. 2, 2021.

[6] COUTINHO, R. M.; OLIVEIRA, B. M.; MARTINS, L. M. “Hospital mortality and healthcare burden during the COVID-19 pandemic in Brazil.” *PLOS ONE*, vol. 16, no. 7, 2021.

[7] SECRETARIA DE ESTADO DA SAÚDE DE SANTA CATARINA. *Relatório Anual de Gestão*. Florianópolis, 2023.

[8] VAN DER LAAN, M. J.; LEE, H.; HUBBARD, A. E. “Discovering patient subgroups with unsupervised learning in large healthcare datasets.” *BMC Medical Informatics and Decision Making*, vol. 21, 2021.

[9] MELO, A. R.; BARROS, C. A. “Unsupervised learning to identify hospital care patterns in Brazilian administrative data.” *Revista Brasileira de Informática em Saúde*, vol. 12, no. 2, 2022.

[10] BARROS, L. C.; MARTINS, M.; PEREIRA, F. “Trends in Hospitalizations in Brazil: A 12-Year Time-Series Analysis Using the SIH/SUS Database.” *Epidemiologia e Serviços de Saúde*, vol. 30, no. 2, 2021.

[11] SHICKEL, B.; TAN, J.; BADLE, A.; YANG, X. ”Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record Analysis.” *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1589–1604, 2018.

[12] JUTEL, A.; SANDS, D. ”Electronic Health Records, Administrative Data, and Big Data in Healthcare: A Review of Uses and Limitations.” *Annual Review of Public Health*, vol. 40, pp. 423–438, 2019.

[13] SILVA, M. E.; GONÇALVES, L. M.; NOVAES, M. A. ”Machine Learning Approaches for Predicting Hospital Mortality Using Brazilian Administrative Health Data (SIH/SUS).” *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, p. 295, 2020.

[14] OUSLEY, J.; ROSSEN, L.; ANDERSON, R. N. ”Forecasting Hospital Utilization Trends Using Time-Series Models Applied to U.S. Administrative Claims Data (Medicare).” *Health Services Research*, vol. 57, no. 3, pp. 583–595, 2022.

[15] BATE, A.; HUTCHINGS, A.; DOMINICZAK, A. ”Unsupervised Learning for Identifying Patient Subgroups in Large NHS Administrative Datasets.” *BMJ Open*, vol. 10, no. 8, 2020.

[16] VERMA, A. A.; KO, D. T.; GUO, H. ”Temporal Trends and Predictive Modeling Using Canadian Administrative Hospital Data: Insights from the CIHI Discharge Abstract Database.” *CMAJ Open*, vol. 10, no. 1, 2022.