Machine Learning-Based Prediction of Inpatient Length of Stay at CNRFR – Rehazenter Rehabilitation Center

Luis Dierick and Benoît Frénay

University of Namur

1 Introduction and Motivation

Length of Stay (LoS) is a central metric in healthcare, reflecting both the clinical progress of patients and the efficiency of resource allocation. While LoS prediction has been widely studied in acute hospital care, its application in rehabilitation settings remains limited. Exploring LoS is particularly important in rehabilitation, where stays are typically much longer than in acute hospital care, making efficient management crucial to optimize resources and improve patient flow.

This study was conducted at the Centre National de Rééducation Fonctionnelle et Réadaptation (CNRFR – Rehazenter) in Luxembourg, a specialized rehabilitation center caring for neurological and traumatological patients. The motivation was twofold: to explore the feasibility of applying Machine Learning (ML) methods for LoS prediction in rehabilitation and to evaluate their potential impact on clinical decision-making and operational management at the Rehazenter.

2 Methodology

The study relied on anonymized data from 1,674 rehabilitation stays, including 667 neurological and 1,007 traumatological cases. The dataset comprised demographic, clinical, and functional features such as age, sex, comorbidities, diagnosis codes from International Classification of Diseases (ICD), Functional Independence Measure (FIM) scores, and other rehabilitation-specific variables.

To address the strong right-skewness of LoS, log transformations and sample weighting strategies were applied. A range of regression algorithms were compared, including Multiple Linear Regression (MLR) and tree-based models such as Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Categorical Boosting (CatBoost) and Cubist. Models were selected with a 5-fold grid search methodology using R^2. To ensure interpretability, SHapley Additive exPlanations (SHAP) were used to identify the most influential features in the predictions.

3 Results

The results show that tree-based ensemble methods clearly outperformed linear models in both cohorts. The best results were achieved with CatBoost and RF, with R^2 values of 0.51 in the traumatological group and 0.25 in the neurological group, while MLR consistently underperformed. The most influential predictors of LoS were FIM scores at admission and the number of chronic diseases. In neurological patients, some comorbidities, socio-economic and psychosocial factors were particularly important, while in traumatological patients, only clinical features were significant. SHAP analysis reinforced these findings by providing transparent explanations of the predictions, highlighting the main drivers of short or long stays.

4 Implications

These findings demonstrate the feasibility of applying ML for LoS prediction in rehabilitation and highlight the potential of such models for integration into clinical workflows. For clinicians, predictive models could support discharge planning and enhance communication with patients and their families by providing realistic expectations regarding the expected duration of stays. For managers, they could offer valuable support for optimizing bed management and allocating resources more effectively, particularly, for the waiting list of the CNRFR – Rehazenter and optimize the bed occupancy rate, which was 93% in 2024. The findings allows also managers to support therapists in organizing their workflows by helping them anticipate patient trajectories and adjust rehabilitation programs accordingly. Beyond the institutional level, such optimizations can also contribute to reducing healthcare expenditures borne by the Caisse Nationale de Santé (CNS), thereby supporting sustainability at the national level.

5 Conclusion and Future Work

This thesis shows that ML-based LoS prediction in rehabilitation is both feasible and valuable, although predictive power remains modest given the complexity of rehabilitation care. The work also underscores the importance of high-quality and structured data for improving model accuracy and ensuring generalizability.

Future work should aim to expand the dataset through multi-center collaborations and larger cohorts, improve the granularity of information by integrating temporal FIM trajectories and psychosocial variables, and investigate the potential of more advanced models such as Neural Network (NN) or hybrid approaches. Finally, developing decision support tools that embed these predictive models into clinical practice could significantly enhance both patient care and the efficiency of healthcare management.

References

[1] Guilherme Almeida et al. "Hospital Length-of-Stay Prediction Using Machine Learning Algorithms—A Literature Review". In: *Applied Sciences* 14.22 (2024), p. 10523.

[2] Belal S. Alsinglawi et al. "Predicting Hospital Stay Length Using Explainable Machine Learning". In: *IEEE Access* 12 (2024), pp. 90571–90585. DOI: 10.1109/ACCESS.2024.3421295.

[3] Ayce Atalay and Nur Turhan. "Determinants of length of stay in stroke patients: a geriatric rehabilitation unit experience". In: *International Journal of Rehabilitation Research* 32.1 (2009), pp. 48–52.

[4] "ATS Statement". In: *American Journal of Respiratory and Critical Care Medicine* 166.1 (2002). PMID: 12091180, pp. 111–117. DOI: 10.1164/ajrccm.166.1.at1102. eprint: https://doi.org/10.1164/ajrccm.166.1.at1102. URL: https://doi.org/10.1164/ajrccm.166.1.at1102.

[5] A. Author. "Sample Article". In: *Journal of Examples* 1.1 (2023), pp. 1–10.

[6] Karen J Brasel et al. "Length of stay: an appropriate quality measure?" In: *Archives of Surgery* 142.5 (2007), pp. 461–466.

[7] S.M Campbell, M.O Roland, and S.A Buetow. "Defining quality of care". In: *Social Science & Medicine* 51.11 (2000), pp. 1611–1625. ISSN: 0277-9536. DOI: https://doi.org/10.1016/S0277-9536(00)00057-5. URL: https://www.sciencedirect.com/science/article/pii/S0277953600000575.

[8] B Catharine Craven et al. "Predicting rehabilitation length of stay in Canada: it's not just about impairment". In: *The journal of spinal cord medicine* 40.6 (2017), pp. 676–686.

[9] Mary E. Charlson et al. "A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation". In: *Journal of Chronic Diseases* 40.5 (1987), pp. 373–383. ISSN: 0021-9681. DOI: https://doi.org/10.1016/0021-9681(87)90171-8. URL: https://www.sciencedirect.com/science/article/pii/0021968187901718.

[10] Davide Chicco, Matthijs J. Warrens, and Giuseppe Jurman. "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation". In: *PeerJ Computer Science* 7 (July 2021), e623. ISSN: 2376-5992. DOI: 10.7717/peerj-cs.623. URL: https://doi.org/10.7717/peerj-cs.623.

[11] Krzysztof J. Cios and G. William Moore. "Uniqueness of medical data mining". In: *Artificial Intelligence in Medicine* 26.1 (2002). Medical Data Mining and Knowledge Discovery, pp. 1–24. ISSN: 0933-3657. DOI: https://doi.org/10.1016/S0933-3657(02)00049-0. URL: https://www.sciencedirect.com/science/article/pii/S0933365702000490.

[12] Aleen Clarke and Rebecca Rosen. "Length of stay: how short should hospital care be?" In: *The European Journal of Public Health* 11.2 (2001), pp. 166–170.

[13] Archibald Danquah-Amoah and Amit Alexander Charan. "The relevance of rehabilitation centers in our communities". In: *International Journal of Science and Research (IJSR)* 6.7 (2017), pp. 1393–1395.

[14] *Document à propos de la FIM du Gouvernement Fédéral Belge*. https://www.health.belgium.be/sites/default/files/uploads/fields/fpshealth_theme_file/fimtekst_fr.pdf.

[15] Molla S Donaldson, Janet M Corrigan, and Linda T Kohn. "To err is human: building a safer health system". In: (2000).

[16] Anne Elixhauser et al. “Comorbidity measures for use with administrative data”. In: *Medical care* 36.1 (1998), pp. 8–27.

[17] Sedigheh Eslami, Peter Adorjan, and Christoph Meinel. “SehMIC: Semi-hierarchical Multi-label ICD code Classification”. In: Sept. 2020.

[18] NHS Executive and Performance Assessment Framework. *Quality and performance in the NHS: high level performance indicators*. NHS Executive, 1999.

[19] Joel S Feigenson et al. “Factors influencing outcome and length of stay in a stroke rehabilitation unit. Part 1. Analysis of 248 unscreened patients–medical and functional prognostic indicators.” In: *Stroke* 8.6 (1977), pp. 651–656.

[20] Alessandro Gasparini. “comorbidity: An R package for computing comorbidity scores”. In: *Journal of Open Source Software* 3 (23 2018), p. 648. DOI: 10.21105/joss.00648. URL: https://doi.org/10.21105/joss.00648.

[21] Deepika Gopukumar and Xiaobo Quan. “Use of claims data to predict the inpatient length of stay among US stroke patients”. In: (2023).

[22] Sneha Grampurohit and Sagar Sunkad. “Hospital Length of Stay Prediction using Regression Models”. In: *2020 IEEE International Conference for Innovation in Technology (INOCON)*. 2020, pp. 1–5. DOI: 10.1109/INOCON50539.2020.9298294.

[23] Dianne L Groll et al. “The development of a comorbidity index with physical function as the outcome”. In: *Journal of clinical epidemiology* 58.6 (2005), pp. 595–602.

[24] Avi Gupta, Thor S. Stead, and Latha Ganti. “Determining a Meaningful R-squared Value in Clinical Medicine”. In: *Academic Medicine & Surgery* (2024). DOI: 10.62186/001c.125154.

[25] D.H. Gustafson. “Length of stay: prediction and explanation”. In: *Health Services Research* 3.1 (1968), pp. 12–34.

[26] *hccinfhir python package*. https://github.com/mimilabs/hccinfhir.

[27] Agency for Healthcare Research and Quality. *Chronic Condition Indicator Refined (CCIR) for ICD-10-CM*. https://hcup-us.ahrq.gov/toolssoftware/chronic_icd10/chronic_icd10.jsp.

[28] Holly Hedegaard et al. “The 2020 International Classification of Diseases, 10th Revision, Clinical Modification injury diagnosis framework for categorizing injuries by body region and nature of injury”. In: (2020).

[29] SNOMED International. *Powering clinical data analytics with SNOMED CT*. https://www.youtube.com/watch?v=hmB3VMu_74w.

[30] SNOMED International. *SNOMED CT*. https://www.snomed.org.

[31] Raunak Jain et al. “Machine Learning Models To Predict Length Of Stay In Hospitals”. In: *2022 IEEE 10th International Conference on Healthcare Informatics (ICHI)*. 2022, pp. 545–546. DOI: 10.1109/ICHI54592.2022.00105.

[32] Raunak Jain et al. “Predicting hospital length of stay using machine learning on a large open health dataset”. In: *BMC Health Services Research* 24.1 (2024), p. 860.

[33] Chihiro Kato et al. “Functional outcome prediction after spinal cord injury using ensemble machine learning”. In: *Archives of physical medicine and rehabilitation* 105.1 (2024), pp. 95–100.

[34] Sascha Kraus et al. “Digital transformation in healthcare: Analyzing the current state-of-research”. In: *Journal of Business Research* 123 (2021), pp. 557–567. ISSN: 0148-2963. DOI: https://doi.org/10.1016/j.jbusres.2020.10.030. URL: https://www.sciencedirect.com/science/article/pii/S0148296320306913.

[35] Bartosz Krawczyk. “Learning from imbalanced data: open challenges and future directions”. In: *Progress in artificial intelligence* 5.4 (2016), pp. 221–232.

[36] Elena Kulinskaya, Diana Kornbrot, and Haiyan Gao. "Length of stay as a performance indicator: robust statistical methodology". In: *IMA Journal of Management Mathematics* 16.4 (2005), pp. 369–381.
[37] Judith R. Lave and Samuel Leinhardt. "The Cost and Length of a Hospital Stay". In: *Inquiry* 13.4 (1976), pp. 327–343. ISSN: 00469580, 19457243. URL: http://www.jstor.org/stable/29771026 (visited on 04/25/2025).
[38] Fung Mei Ling. "Stroke rehabilitation: predicting inpatient length of stay and discharge placement". In: *Hong Kong Journal of Occupational Therapy* 14.1 (2004), pp. 3–11.
[39] Scott M Lundberg and Su-In Lee. "A unified approach to interpreting model predictions". In: *Advances in neural information processing systems* 30 (2017).
[40] Willard G Manning and John Mullahy. "Estimating log models: to transform or not to transform?" In: *Journal of health economics* 20.4 (2001), pp. 461–494.
[41] Centers for Medicare & Medicaid Services. *MS-DRG Classifications and Software.* https://www.cms.gov/medicare/payment/prospective-payment-systems/acute-inpatient-pps/ms-drg-classifications-and-software.
[42] Centers for Medicare & Medicaid Services. *Risk Adjustment 2025 Model Software ICD-10 Mappings.* https://www.cms.gov/medicare/payment/medicare-advantage-rates-statistics/risk-adjustment/2025-model-software/icd-10-mappings.
[43] National Library of Medicine. *Medical Subject Headings (MeSH).* https://www.ncbi.nlm.nih.gov/mesh/.
[44] Diane Podsiadlo and Sandra Richardson. "The timed "Up & Go": a test of basic functional mobility for frail elderly persons". In: *Journal of the American geriatrics Society* 39.2 (1991), pp. 142–148.
[45] Hude Quan et al. "Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data". In: *Medical care* 43.11 (2005), pp. 1130–1139.
[46] John R Quinlan et al. "Learning with continuous classes". In: *5th Australian joint conference on artificial intelligence*. Vol. 92. World Scientific. 1992, pp. 343–348.
[47] Rehabilitation indicator menu: a tool accompanying the Framework for Rehabilitation Monitoring and second edition. Geneva: World Health Organization; 2023. Licence: CC BY-NC-SA 3.0 IGO. Evaluation (FRAME).
[48] *Rehazenter annual report 2023.* https://www.rehazenter.lu/wp-content/uploads/2024/09/Rapport-Annuel-RHZ-2023_compressed.pdf.
[49] *Rehazenter Missions.* https://www.rehazenter.lu/fr/a-propos/presentation/.
[50] G.H. Robinson, L.E. Davis, and R.P. Leifer. "Prediction of hospital length of stay". In: *Health Services Research* 1.3 (1966), pp. 287–300.
[51] Valter Santilli et al. "The use of machine learning for inferencing the effectiveness of a rehabilitation program for orthopedic and neurological patients". In: *International Journal of Environmental Research and Public Health* 20.8 (2023), p. 5575.
[52] Mark A Schuster, Elizabeth A McGlynn, and Robert H Brook. "How good is the quality of health care in the United States?" In: *The Milbank Quarterly* 76.4 (1998), pp. 517–563.
[53] Farah Shamout, Tingting Zhu, and David A Clifton. "Machine learning for clinical outcome prediction". In: *IEEE reviews in Biomedical Engineering* 14 (2020), pp. 116–126.
[54] Kieran Stone et al. "A systematic review of the prediction of hospital length of stay: Towards a unified framework". In: *PLOS Digital Health* 1.4 (2022), e0000017. DOI: 10.1371/journal.pdig.0000017.

[55] Angelos I. Stoumpos, Fotis Kitsios, and Michael A. Talias. “Digital Transformation in Healthcare: Technology Acceptance and Its Applications”. In: *International Journal of Environmental Research and Public Health* 20.4 (2023). ISSN: 1660-4601. DOI: 10.3390/ijerph20043407. URL: https://www.mdpi.com/1660-4601/20/4/3407.

[56] LATANYA SWEENEY. “ACHIEVING k-ANONYMITY PRIVACY PROTECTION USING GENERALIZATION AND SUPPRESSION”. In: *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10.05 (2002), pp. 571–588. DOI: 10.1142/S021848850200165X. eprint: https://doi.org/10.1142/S021848850200165X. URL: https://doi.org/10.1142/S021848850200165X.

[57] LATANYA SWEENEY. “k-ANONYMITY: A MODEL FOR PROTECTING PRIVACY”. In: *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10.05 (2002), pp. 557–570. DOI: 10.1142/S0218488502001648. eprint: https://doi.org/10.1142/S0218488502001648. URL: https://doi.org/10.1142/S0218488502001648.

[58] Woan Shin Tan et al. “Factors Predicting Inpatient Rehabilitation Length of Stay of Acute Stroke Patients in Singapore”. In: *Archives of Physical Medicine and Rehabilitation* 90.7 (2009), pp. 1202–1207. ISSN: 0003-9993. DOI: https://doi.org/10.1016/j.apmr.2009.01.027. URL: https://www.sciencedirect.com/science/article/pii/S0003999309002731.

[59] Lior Turgeman, Jerrold H May, and Roberta Sciulli. “Insights from a machine learning model for predicting the hospital Length of Stay (LOS) at the time of admission”. In: *Expert Systems with Applications* 78 (2017), pp. 376–385.

[60] *What is machine learning?* https://www.ibm.com/think/topics/machine-learning.

[61] Tara A Whitten et al. “Predicting inpatient rehabilitation length of stay for adults with traumatic spinal cord injury”. In: *The Journal of Spinal Cord Medicine* (2024), pp. 1–11.

[62] *WHO classifications.* https://www.who.int/standards/classifications.

[63] *WHO ICD-10 Manual.* https://icd.who.int/browse10/Content/statichtml/ICD10Volume2_en_2019.pdf.

[64] Zhenhui Xu et al. “Predicting in-hospital length of stay: a two-stage modeling approach to account for highly skewed data”. In: *BMC Medical Informatics and Decision Making* 22.1 (2022), p. 110.

[65] So-Mei Teresa Yeung, Aileen M Davis, and Rajka Soric. “Factors influencing inpatient rehabilitation length of stay following revision hip replacements: a retrospective study”. In: *BMC Musculoskeletal disorders* 11 (2010), pp. 1–10.