PROGNURSE

Projecting Nursing Workload using Routine Data Analytics - A Machine Learning Approach

Graduate

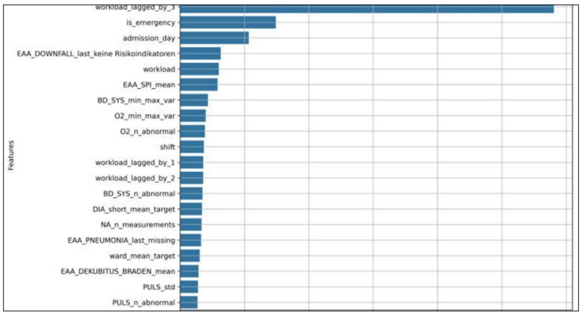
Samuel Kolb

Problem: Hospitals and other care providers in Europe are experiencing a significant nursing shortage, leading to an increased nursing workload. This rise in workload is associated with higher burnout rates and reduced job satisfaction among nurses, potentially causing them to leave the profession or seek other roles. Moreover, the increased workload adversely impacts patient safety and satisfaction by disrupting care continuity and overall patient-centered care. To address this issue, early information about the expected workload in upcoming shifts per patient is essential for appropriate staffing.

Approach: This master's thesis, in collaboration with the University Hospital Basel (USB), investigates the feasibility of predicting the nursing workload for the upcoming three shifts based on historical workloads, originating from LEP (Leistungserfassung in der Pflege), patient characteristics, ward information, and various routine patient data, including diagnoses, admission type, vital values, laboratory parameters, and assessment scores. The study explores and implements suitable machine learning models to learn patterns from past patient cases and predict nursing workload for new cases starting from the first admission timestep. Specifically, machine learning-based global forecasting models are examined, including classic regression models such as Linear Regression, Random Forest, and XGBoost, as well as deep learning-based regression models such as LSTM, TFT, and TSMixer. The necessary data preprocessing steps to apply these models to the patient cases data are detailed, and the results are evaluated using MAE and R^2 on unseen patient cases data. Feature importance analyses identify critical predictive variables.

Result: The results indicate that the top-performing models achieve a Mean Absolute Error (MAE) of 35 to 37 minutes over the prediction horizon, explaining 55% to 60% of the variance in unseen patient cases. While the linear model is outperformed by more complex models, the performance is comparable across tree-based and deep learning models. Notably, predictions improve significantly from the second admission timestep onward, although the MAE tends to increase as the prediction horizon extends. Most workloads yield promising results, but the models struggle with rare, exceptionally high workloads. At the ward level, aggregated predictions show that certain wards allow for more accurate forecasts, suggesting that initial estimates of the required nursing staff can be effectively made using these models for those wards.
The results of applying these models in hospital settings demonstrate promising generalizability. Additionally, the author, in collaboration with experts at USB, has identified potential for a follow-up research project. This new project will aim to extend the prediction horizon and integrate nursing qualifications to simulate optimal scenarios for nursing staff planning, thereby reducing workload peaks on wards. In light of these findings, specific recommendations for future research are proposed.

Feature importance of the TFT encoder variables.
Source: Author



Advisor
Dr. Shao Jü Woo

Co-Examiner
Mag. Dipl.-Ing. Dr. Kathrin Plankensteiner

Subject Area
Data Science

Project Partner
University Hospital Basel, Basel, Basel-Stadt