



HAL
open science

Nurses' workload prediction in hospitals -a Machine Learning-based approach

Mohamed Gharbi, Christine Verdier, Maria Di Mascolo

► **To cite this version:**

Mohamed Gharbi, Christine Verdier, Maria Di Mascolo. Nurses' workload prediction in hospitals -a Machine Learning-based approach. INFORSID, May 2024, Nancy Université de Lorraine, France. hal-04610859

HAL Id: hal-04610859

<https://hal.science/hal-04610859>

Submitted on 13 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Nurses' workload prediction in hospitals – a Machine Learning-based approach

Mohamed GHARBI

Calystene, 3C rue Irène Joliot Curie, 38320 Eybens, France.

Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France.*

Univ. Grenoble Alpes, CNRS, Grenoble INP, G-SCOP, 38000, Grenoble, France.*

** Institute of Engineering Univ. Grenoble Alpes*

mohamed.gharbi1@univ-grenoble-alpes.fr

ABSTRACT. In hospitals, nurses represent nearly half of the staff and they play an important role in healthcare delivery. However, there is an existing shortage that leads to increased workloads, which affects their performances and the quality of care they provide. To address this issue, workload prediction has been suggested as a solution. Despite its potential benefits, workload prediction presents challenges such as data selection, data size, and model selection. In this article, we will examine various methods and models utilized for predicting nurses' workload, as well as in other areas such as energy. Additionally, we will propose a methodological approach to overcome the limitations associated with workload prediction in our future works.

KEYWORDS. Workload, data processing, Machine Learning.

SUPERVISION. Christine Verdier (LIG), Maria Di Mascolo (G-SCOP).

1. Introduction

In contemporary healthcare, predicting nurses' workload has emerged as one of the top priorities for hospitals seeking to ensure optimal patient care. Despite the crucial role that nurses play in healthcare delivery, there is a significant shortage of the nursing-staff, which is affecting the hospitals ability to meet the evolving healthcare needs. To address this challenge, considerable efforts have been dedicated to understand and explore the factors influencing nurses' workload and to develop predictive models supporting nursing work management.

Although predicting workload for nurses' remains an ongoing attempt, progress has been made in several fields, including human resources, and energy. In Data Science, numerous Machine Learning models such as linear regression, sequence models like recurrent neural networks (RNNs), especially Long Short-Term Memory

(LSTM) and probabilistic models have been applied to tackle this issue. These models can handle different data types like sparse, heterogeneous, and big data.

In this paper, we describe and develop our global architecture in the field of nurses' workload prediction by outlining the possible paths for our data and models preparation and selection.

2. State of the art

Numerous studies have explored methods for calculating nurses' workload and identifying the factors that influence it within hospital settings. In our prior research (Gharbi *et al.*, 2024), we reviewed these methodologies and outlined the envisioned advancements in the area. In addition, we highlighted the several approaches that were used for the nurses' workload prediction and their limits in combining and using the influential factors such as patients characteristics, nurses experience, the complexity of cares and their variability status. Some studies have highlighted the important role of Machine Learning in predicting nurses' workload; for instance, addressing flexible staffing issues (Kortbeek *et al.*, 2015).

Machine learning has already proven beneficial across other domains of healthcare industry, like in clinical trials ranging from preclinical drug discovery to pre-trial planning, facilitated by robust data management and analysis (Weissler *et al.*, 2021). Additionally, (Mirza *et al.*, 2023) focused on source data verification (SDV) prediction using LSTM and time series analysis. Furthermore, (Mughees *et al.*, 2021) applied Bi-LSTM for day-ahead peak load forecasting within the power industry. While these Machine Learning techniques are adopted to manage extensive and diverse datasets, (Alsafadi and Wu, 2023) emphasized the utility of Generative Deep Models in tackling challenges associated with smaller datasets. They also discussed the application of random forest models for feature capture.

Nonetheless, when it comes to predicting nurses' workload – due to its complex nature influenced by the cited factors and the different data types– the application of machine learning in this area remains a field for improvement.

3. Problematic

In our research, we encounter various challenges and limitations related to understanding and processing the data, especially considering its size, variation, and complexity. Additionally, selecting of the appropriate models can be challenging based on the expected outcomes. **Data understanding** phase has the goal of identifying and validating the key factors influencing nurses' workload, drawing from their professional experience. This necessitates engaging a significant number of nursing staff to gather meaningful insights. As for **data processing**, it presents its own set of challenges. Our dataset is heterogeneous, consisting of diverse types of data such as numeric, qualitative and textual, in addition, several data sources (medical record, nurses' schedules, PRN etc.). The task is to effectively integrate and leverage this multifaceted data. A notable issue we face is the limited size and sequential nature

of our dataset. This causes some difficulties when considering the application of deep learning models, such as LSTM. While these models have demonstrated effectiveness in prior research, as shown by (Mirza *et al.*, 2023), they typically require extensive data, which in their study spanned from 2014 to 2020.

This leads us to several **pertinent questions**: Where can we access a more extensive dataset? Is a large dataset indispensable for our objective? Could an alternative approach involving the generation of sequential data, while fine-tuned using methods such as Hidden Markov models (Hanif *et al.*, 2017), be a viable solution? Additionally, we need to address how to manage missing data effectively while preserving as much valuable information as possible.

4. Contribution

Based on the literature and the perspectives mentioned in the article (Gharbi *et al.*, 2024), data understanding, handling and the use of predictive models are considered as potential contributions. As a follow up, we will discuss these contributions in detail. We propose a structured approach that encompasses the various steps of our future contributions, drawing from a standard data-mining model derived from the CRISP-DM framework (Mirza *et al.*, 2023; Wirth and Hipp, 2000). In this approach, we will outline our planned work for the data processing and data modelling stages.

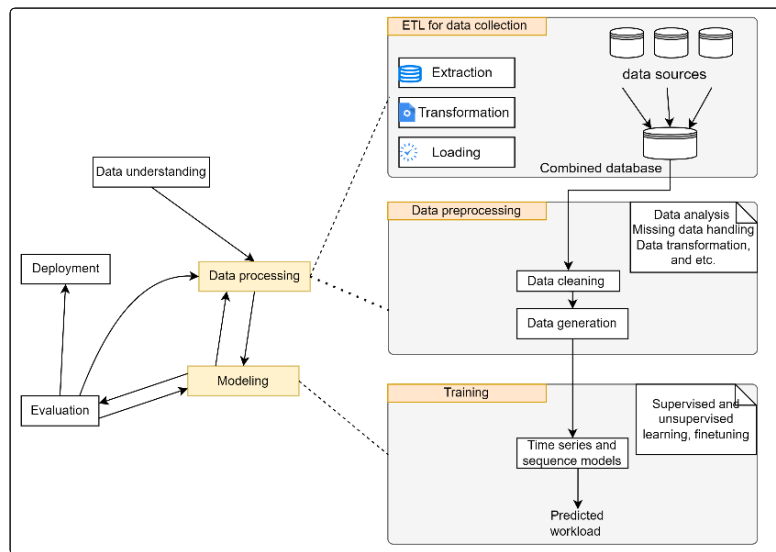


Figure 1. Nurses' workload prediction process: from data understanding to deployment

Figure 1, illustrates the process which begins with **data understanding**; this initial phase enables us to grasp our data, identify relevant variables such as patient and

nurses' characteristics, care types, and specific details about wards and hospitals. The second phase is **data processing** that is divided into two steps. First, we collect data from different sources, including the computerized patient record for patient related data and HR systems for staff-related data, ensuring anonymization for ethical and security reasons. These different sources are consolidated into a unified database following the ETL (extract, transform, and load) process (Hendayun *et al.*, 2021). Second, we analyse and clean our data by addressing missing values, examining correlations between variables, and handling unbalanced classes. Given the small size of our dataset, we aim to explore the potential of generating sequential data inspired by unsupervised learning especially the Generative Deep Models. Finally, techniques such as normalization, outlier detection, and statistics might be employed. For the **prediction model**, entry data might be processed and transformed using machine learning model such as time-series. Subsequently, a ML model that deals with sequential data, can be applied. We then predict workload by ward, expressed in hours across different time slots based on the ward organization (eg. morning, afternoon, and night). Following this, we evaluate our model and deploy it. The result is a dashboard that exploits the several key performance indicators.

As **future work**, we plan to refine our methodological approach on which we will have a process that generates meaningful data and train models that address our challenges. We plan to study and employ Generative Deep Models (GDMs), Long Short-Term Memory (LSTMs) networks, and Time-Series analysis. Additionally, we are interested in exploring the potential of incorporating spatial data into our research.

References

- Alsafadi, F., Wu, X., (2023). *Deep generative modeling-based data augmentation with demonstration using the BFBT benchmark void fraction datasets*. Nucl. Eng. Des. 415, 112712.
- Gharbi, M., Di Mascolo, M., Verdier, C. (2024). Charge de travail du personnel infirmier dans les hôpitaux -étude bibliographique. Congrès conjoint Alass-Giseh 2024. 3-6 juillet 2024. Liège. Belgique
- Hendayun, M., Yulianto, E., Rusdi, J.F., Setiawan, A., Ilman, B., (2021). *Extract transform load process in banking reporting system*. MethodsX 8, 101260.
- Mirza, B., Li, X., Lauwers, K., Reddy, B., Muller, A., Wozniak, C., Djali, S., (2023). *A clinical site workload prediction model with machine learning lifecycle*. Healthc. Anal. 3, 100159.
- Mughees, N., Mohsin, S.A., Mughees, Abdullah, Mughees, Anam, (2021). *Deep sequence to sequence Bi-LSTM neural networks for day-ahead peak load forecasting*. Expert Syst. Appl. 175, 114844.
- Weissler, E.H., Naumann, T., Andersson, T., Ranganath, R., Elemento, O., Luo, Y., Freitag, D.F., Benoit, J., Hughes, M.C., Khan, F., Slater, P., Shameer, K., Roe, M., Hutchison, E., Kollins, S.H., Broedl, U., Meng, Z., Wong, J.L., Curtis, L., Huang, E., Ghassemi, M., (2021). *The role of machine learning in clinical research: transforming the future of evidence generation*. Trials 22, 537.