Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240587

Evaluation of Machine Learning Algorithms for Pressure Injury Risk Assessment in a Hospital with Limited IT Resources

Cynthia ABI KHALIL[a,b,1], Antoine SAAB[a,c], Jihane RAHME[b,c], Jana ABLA[b] and Brigitte SEROUSSI[a,d]
[a]*Sorbonne Université, Université Sorbonne Paris Nord, INSERM, Laboratoire d'Informatique Médicale et d'Ingénierie des connaissances en e-Santé, LIMICS, F-75006 Paris, France*
[b]*Nursing Administration, Lebanese Hospital Geitaoui–UMC, Beirut, Lebanon*
[c]*Quality and Safety Department, Lebanese Hospital Geitaoui–UMC, Beirut, Lebanon*
[d]*AP-HP, Hôpital Tenon, Paris, France*
ORCiD ID: Cynthia Abi Khalil https://orcid.org/0000-0001-9112-628X
Antoine Saab https://orcid.org/0000-0003-0673-0875
Brigitte Séroussi https://orcid.org/0000-0002-7676-2832

Abstract. Clinical decision support systems for Nursing Process (NP-CDSSs) help resolve a critical challenge in nursing decision-making through automating the Nursing Process. NP-CDSSs are more effective when they are linked to Electronic Medical Record (EMR) Data allowing for the computation of Risk Assessment Scores. Braden scale (BS) is a well-known scale used to identify the risk of Hospital-Acquired Pressure Injuries (HAPIs). While BS is widely used, its specificity for identifying high-risk patients is limited. This study develops and evaluates a Machine Learning (ML) model to predict the HAPI risk, leveraging EMR readily available data. Various ML algorithms demonstrated superior performance compared to BS (pooled model AUC/F1-score of 0.85/0.8 vs. AUC of 0.63 for BS). Integrating ML into NP-CDSSs holds promise for enhancing nursing assessments and automating risk analyses even in hospitals with limited IT resources, aiming for better patient safety.

Keywords. Clinical Decision Support Systems, Nursing Process, Assessment, Artificial Intelligence, Machine Learning, Hospital acquired Pressure Injury

1. Introduction

Clinical decision support systems (CDSSs) that support the Nursing Process (NP) are systems that enhance nursing decision-making throughout a comprehensive approach of the Nursing Process: assessment, diagnosis, planning, implementation and evaluation [1]. Whereas NP-CDSSs are considered as holding promises for improving nursing quality of care, recommendations from the 2016 internationally consented standard for NP-

[1] Corresponding Author: Cynthia ABI KHALIL; E-mail: cynthiakjammal@outlook.com.

CDSSs [1] highlighted the importance of automated linkages between the five NP phases and the integration with other components of the EMR, in particular with Risk Assessment Scores. The most famous risk assessment scores in nursing are those used to identify Hospital Acquired Pressure Injuries (HAPIs) defined as damages localized on the skin and underlying soft tissues, usually over a bony prominence or related to a medical device [2].

Considered as costly but preventable adverse events, HAPIs are one of the most important nursing-specific quality indicators and a crucial nursing diagnosis [2] for which patients' complications may be reduced when risk is identified early and preventive measures are implemented in a timely manner [3]. The Braden scale (BS) is commonly used to assess the risk for HAPI. BS incorporates information from six subscales (score between 6-23, with lower scores indicating more severe risk) [2]. Nevertheless, BS may not be useful as it may result in unnecessary expenditure of time and resources. The reported low specificity of BS for identifying high risk patients hence begs the inclusion of other important predictors. This has led to the expansion and evaluation of covariates sought to predict HAPI incidence [2]. Machine learning (ML) models using EMR data to predict the risk of HAPI are starting to show evidence in assisting nurses to identify HAPI earlier in time and with a high accuracy rate, thus promoting patient safety [3].

The objective of this work is to develop a ML model integrated to a NP-CDSS to predict patients at high risk of HAPI based on variables highlighted by the literature and sought from a basic EMR implemented in a hospital with limited IT resources.

2. Materials and Methods

We followed the MINIMAR (Minimum Information for Medical AI Reporting) standards for reporting the study results [4].

2.1. Study population, settings, and patient demographics

The study was conducted in a 250-bed tertiary university hospital in Beirut, Lebanon. The hospital's EMR (HIMSS EMRAM stage 2 [5]) includes admissions/discharge/transfer data, basic ancillaries with limited integration (i.e., laboratory, radiology, and pharmacy), billing information (i.e., procedures and consumables), but no electronic nursing or medical documentation, and no computerized physician order entry or CDS applications. We identified all consecutive inpatients from 1 January 2017 through 7 March 2024, in all hospital wards (medical, surgical, and critical care), with length of stay of more than 48 hours, and for which a HAPI prevention mattress and/or HAPI dressing has been billed, indicating patients had been considered at high risk for HAPI, yielding a cohort of 1,087 cases. From the remaining set of patients (considered as low risk for HAPI) admitted as inpatients in the same time period and length of stay condition, 999 cases were randomly selected and added to the first cohort, yielding to a resulting dataset of 2,086 cases considered as balanced for HAPI risk. Eighty percent of the dataset was selected through stratified randomization to maintain balance in risk as the training set (1,669 cases), and the remaining 417 cases were used as the testing set to assess the model performance. The demographic characteristics of the training and testing datasets are shown in Table 1.

Table 1. Demographic characteristics of the training and testing datasets.

	Age Avg (Std Dev)	Gender F/M	Length of stay Avg (Std Dev)	HAPI risk #High/#Low
Training data set	70.7 (21.7)	837/832	6.7 (8.1)	855/814
Testing data set	72.1 (20.7)	209/208	7.0 (7.0)	232/185
Total	71.0 (21.6)	1,046/1,040	6.8 (7.9)	1087/999

Avg: Average, Std Dev: Standard Deviation, #: Count

2.2. Model Architecture

A supervised learning approach was adopted based on the different ML algorithms used in similar studies [2,3]: random forests (RF), Decision Tree (DT), gradient boosting (XGBoost), and logistic regression (LR). Python programming language was used for developing the scripts to create and analyze ML models (Sklearn library [6]).

Thirty-four features were identified as relevant based on domain knowledge and existing literature [2,3,7,8]. These features include: demographic variables (e.g., age, gender, length of stay), comorbidities (eg., polypharmacy index), nutritional status (e.g., parenteral and enteral feeding), laboratory results (e.g., albumin level, C-Reactive Protein (CRP), Hemoglobin). We preprocessed data by handling outliers and reporting missing values per feature.

For each ML algorithm, the output is the categorical classification of the HAPI risk level (High/Low) for a patient, as a decision support for nurses, based on available, although limited, routine data in the EMR.

2.3. Model Evaluation

For each ML algorithm, parameters were trained and tuned through a 5-fold cross validation on the training dataset (not the testing set).

The area under the curve (AUC) and the F1-score (harmonic mean of the precision and recall scores*)* were used to assess the performance of the different algorithms relative to the cross-validation of training and testing datasets. Additionally, we considered the Braden scale retrieved on the paper patient record as the gold standard for the HAPI risk assessment. BS was identified through a retrospective chart review of patient cases in the testing dataset by two experienced nurses (JR & JA) and reported as AUC for three threshold values of BS (scores 12, 15&18).

To enhance clinical interpretability and identify the important predictors of the models, each explanatory variable importance was assessed by their relative influence on the algorithm classification results using the Python Sklearn library [6].

3. Results and Discussion

Results of the different algorithms' performance (AUC/F1-Score) and the comparison on the testing dataset with the three thresholds for BS are reported in Table 2. Variable importance scores computed from the XGBoost algorithm (which showed the highest performance) are displayed in Figure 1.

3.1. Performance of the different predictive models and explanatory variables

ML models achieved a good performance across the tested algorithms with a pooled AUC and F1-score respectively of 0.85 and 0.80. The lack of EMR features from clinical documentation was partially compensated by the choice of proxy variables stemming from billing codes that indicate clinical procedures and patient status using readily available data without the need for input from clinical professionals. Age, length of stay, number of comorbidities were the most predictive variables, followed by Alkaline Reserve, CRP, Hemoglobin, Hematocrit and Albumin level. This result is well in line with a recent review on the performance of ML algorithms for HAPI prediction [3], where AUC of the models ranged from 0.74 to 0.94, and most prevalent variables used were age, gender, body mass index, length of stay, medications, vital signs, anesthesia, BS, diagnoses including cancer, cardiovascular, diabetes mellitus, renal failure, respiratory, and diagnostic tests.

Table 2. Comparison of models' performance.

Algorithm	Model Validation AUC	Model Testing AUC	Model Testing F1-score	Braden scale AUC (Cut=12)	Braden scale AUC (Cut=15)	Braden scale AUC (Cut=18)
RF	0.88	0.86	0.79	0.57	0.60	0.63
XGB	0.89	0.88	0.82			
LR	0.80	0.79	0.76			
DT	0.83	0.84	0.80			
RF: Random Forest, XGB: XGBoost, LR: Logistic Regression, DT: Decision Tree, Cut: threshold						

3.2. Comparison with the gold standard (Braden Scale)

Models significantly outperformed the gold standard (BS) on the same testing dataset. BS performance was relatively low (with a maximum AUC of 0.63 for threshold score 18), mainly due to a low specificity estimated at 0.36. This result is coherent with recent reviews relative to the use of ML models for HAPI risk prediction [3,9]. ML higher performance can be related to the use of additional variables relative to patient-specific characteristics (such as age, comorbidity, and other biological markers reflecting low perfusion, oxygenation and high inflammation/edema) that are not included in BS but can still affect the risk of HAPI even when mechanical risk factors are significant.

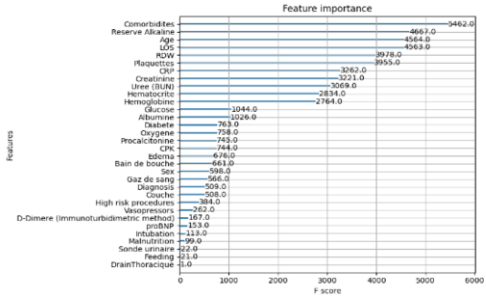

Figure 1. Visualization of variable importance scores (XGBoost algorithm).

3.3. Practical implications for the nursing assessment phase of NP-CDSSs

ML models hold the potential to be integrated in the assessment phase of NP-CDSS tools to automate the risk analysis of HAPI, paving the way for using similar methods to

automating other types of nursing risk assessments and lessening cognitive and documentation burden for nurses. Additionally, explainability features of algorithms may allow the production of a relatively transparent feedback to the user relative to the prediction given for each instance, which may allow nurses to understand recommendations. The good performance of ML models with features stemming from a basic EMR indicates that it can be suitable for systems in low-IT resources settings.

The study has certain limitations. The choice of a proxy measure (HAPI prevention mattress and/or HAPI dressing billings) used to label positive cases could be prone to errors relative to billing procedures. However, it permits easily building of large datasets to train ML models. The other limitation is related to the lack of access to digital clinical notes and biomedical data in EMRs due to the IT low-resource status of the institution.

4. Conclusions

Limitations in the specificity and clinical utility of existing risk assessment scores highlight the need for alternative approaches. ML models, leveraging data from EMRs, show potential in accurately predicting HAPI by incorporating various patient factors beyond those considered in traditional risk scores. Our study explores the potential of integrating ML models into NP-CDSSs for HAPI risk prediction, achieving superior performance as compared to the gold standard (BS assessment). By alleviating the burden of manual risk assessment and providing timely accurate and dynamic predictions, NP-CDSSs equipped with ML models should enhance patient safety and nursing quality of care. While our study acknowledges certain limitations, such as the choice of labeling positive cases, it underscores the potential of leveraging EMR data to train ML models for risk assessment scores and integrate them into clinical workflows in hospitals with limited IT resources.

References

[1] Müller-Staub M, et al. An internationally consented standard for nursing process-clinical decision support systems in electronic health records. CIN - Computers Informatics Nursing. 2016;34(11).
[2] Levy JJ, et al. Machine Learning Approaches for Hospital Acquired Pressure Injuries: A Retrospective Study of Electronic Medical Records. Front Med Technol. 2022;4. DOI: 10.3389/fmedt.2022.926667
[3] Barghouthi ED, Owda AY, Asia M, Owda M. Systematic Review for Risks of Pressure Injury and Prediction Models Using Machine Learning Algorithms. Vol. 13, Diagnostics 2023.
[4] Hernandez-Boussard T, et al. MINIMAR (MINimum information for medical AI reporting): Developing reporting standards for artificial intelligence in health care. Journ of the AMIA 2020;27.
[5] HIMSS. Healthcare Information and Management Systems Society, Inc. [Internet]. 2022. Electronic Medical Record Adoption Model
[6] Pedregosa F, Varoquaux G, et al. Scikit-learn: Machine Learning in Python Pedregosa, Varoquaux, Gramfort et al. Journal of Machine Learning Research. 2011;12.
[7] Wang N, Lv L, Yan F, Ma Y, Miao L, Foon Chung LY, et al. Biomarkers for the early detection of pressure injury: A systematic review and meta-analysis. Vol. 31, Journal of Tissue Viability 2022.
[8] Bazaliński D, et al. Selected biochemical blood parameters and a risk of pressure ulcers in patients receiving treatment in intensive care units. Medicina (Lithuania). 2021;57(2).
[9] Dweekat OY, Lam SS, McGrath L. Machine Learning Techniques, Applications, and Potential Future Opportunities in Pressure Injuries (Bedsores) Management: A Systematic Review. Vol. 20, International Journal of Environmental Research and Public Health 2023.