Interdisciplinary Human-Centered AI for Hospital Readmission Prediction of Heart Failure Patients

Amira SOLIMANa,1, Monika NAIRa, Marcus PETERSSONb, Lina LUNDGRENa, Petra DRYSELIUSB, Ebba FOGELBERGB, Omar HAMEDa, Kobra ETMINANIA and Jens NYGRENA

a Halmstad University, Halmstad, Sweden
b Cambio Healthcare Systems AB, Sweden

Abstract. The evolution of clinical decision support (CDS) tools has been improved by usage of new technologies, yet there is an increased need to develop user-friendly, evidence-based, and expert-curated CDS solutions. In this paper, we show with a use-case how interdisciplinary expertise can be combined to develop CDS tool for hospital readmission prediction of heart failure patients. We also discuss how to make the tool integrated in clinical workflow by understanding end-user needs and have clinicians-in-the-loop during the different development stages.

Keywords. Interdisciplinary Healthcare, Human-Centered AI, Hospital Readmission Prediction.

1. Introduction

Unscheduled rehospitalization of heart failure (HF) patients has received a lot of research attention due to its high cost and undesirable impact on patient outcomes and healthcare planning [1]. Providing a care plan with rapid follow-up and treatment compliance can be effective in minimizing the risk of readmissions [2], however, identifying which patients to prioritize remains to be a challenge. A wide range of predictive modeling methods have been used to address this problem, however, none have been implemented as clinical decision support (CDS) applications in practice using Artificial Intelligence (AI) models that learn and adapt to patterns expressed in patient data.

Sensitive and eXplainable Artificial Intelligence (XAI) models can assist clinicians in managing HF patients at discharge by highlighting individual aspects associated with high-risk of readmission. Previous studies showed the potential applicability in terms of cost savings [3]. However, to gain trust by clinicians, interpretability of model results is needed. Additionally, from a human-centered design perspective, interpretability is not a property of the Machine Learning (ML) model only, but a relationship between the model and the user [4]. Human-centered AI systems operating jointly, rather than alone, have a great potential for high effectiveness [5], however, the design process is complex and requires interdisciplinary as well as interorganizational collaborations.

1 Corresponding Author: Amira Soliman, E-mail: amira.soliman@hh.se.
The objective of this paper is to present a use case developing a CDS application to predict HF patient risk of readmission. The work process had the following parts: 1) identify barriers and enablers for implementation of a CDS application, 2) develop an XAI model, and 3) explore how to present the model output to users.

2. Methods

This research was composed of three interlinked work packages (WP) as shown in Figure 1. The group consisted of university researchers from computer science, UX-design and implementation science, one IT company with expertise in data science and UX-design, and a healthcare organization providing care to HF patients.

2.1. Collecting Stakeholder Input

To ensure relevance of the application to the clinical context, input from stakeholders in Region Halland in Sweden was assembled [6]. Twelve interviews were performed in 2021 with medical specialists in cardiology, specialist nurses, a physiotherapist, a home care physician, a home care nurse and a controller. The semi-structured interviews were conducted one-to-one via video calls, recorded and transcribed verbatim. The questions concerned potential influence the application might have on patients, healthcare staff, and the organization, as well as what could be the barriers and enablers of potential implementation of such a system in practice. The data were thematically coded independently by two researchers [7] using the categories of the NASSS implementation framework [8].

2.2. AI-enabled Readmission Prediction

Retrospective data from the regional healthcare information platform in Region Halland was used for model development [9]. The cohort consisted of patients ≥40 years of age, diagnosed with HF according to ICD-10 (I11.0, I42, I43, and I50), who had at least a single admission after being diagnosed with HF between January 1, 2017, and December 31, 2019. An ML model was developed with CatBoost resembles gradient-boosting decision trees that handles categorical and continuous features [10]. The model makes predictions using a series of decision trees and was trained using k-fold cross-validation for 10 iterations using the training dataset. The SHapley Additive exPlanations (SHAP) technique was adopted to provide more details behind the model decision regarding important features for readmission prediction [11]. SHAP computes a score for each individual prediction to illustrate features that are positively or negatively driving toward
readmission risk. The provided explanations were addressed by physicians for clinical relevance.

2.3. Prototype Design

Based on stakeholder input, low-fidelity sketches were made and tested with three users. After iterating the design, a high-fidelity prototype was tested with three new users. For each usability test, the think-aloud protocol was used followed by a short discussion. All users were practicing or former clinicians with HF expertise.

3. Results

Overall, the reasons for iterations between WPs were; need for additional input data to the model (information-based decisions), stakeholder input on when and how to use the tool (process-based decisions), and stakeholder input on information relevance (interface and utility-based decisions).

3.1. Stakeholder Input

Reducing subjectivity in assessing the risk of readmission for patients was seen as a key value for the clinical context. The stakeholders urged that units outside of cardiology dealing with HF patients, mainly due to a lack of beds, should have access to the CDS application to direct the right patients to the cardiology unit and ensure the right treatment and a reduced risk of future readmission. Free-text medical notes were deemed highly important for the algorithm, as they hold relevant information. The risk assessment should be available early, during admission, thereby enabling contacts with cardiology unit, optimal resource allocation, and prioritization. In addition to risk assessment, the application should present factors that led to such assessments. Lack of interoperability with the existing IT applications in the hospital could hinder adoption. A pilot study, information campaign, training, resource allocation, revised procedures, and competent implementation leaders persons were mentioned as powerful enablers for successful adoption of the application.

3.2. Model Performance and Explainability

The model produced a probability between 0 and 1 to indicate the likelihood of readmission. We used a threshold of value (0.43) to convert the generated probability to classification label, i.e., to separate patients with high readmission risk from others. The threshold value was estimated during model training. Thus, a probability below 0.43 indicates low-risk with label 0, while a probability above 0.43 indicates high-risk of readmission with label 1. Table 1 displays the achieved performance by the model on unseen testing data. Interestingly, the developed model gained better performance compared to both LACE index and HOSPITAL score that are widely used as risk assessment tools to discriminate readmissions using readily available clinical predictors that can be calculated before discharge, yet they are not trainable as our model.

Model explainability is key to clarify the risk. Feature ranking generated by SHAP shows that besides clinically informative variables, administrative features have high importance, e.g., patients with frequent inpatient episodes (Figure 2a). While these features cannot be altered, they are useful warning signals when planning for the discharge.
### Table 1. Results obtained from training, validation, and testing model with data from Region Halland

<table>
<thead>
<tr>
<th></th>
<th>CatBoost Model</th>
<th>LACE Index</th>
<th>HOSPITAL Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>80</td>
<td>40</td>
<td>49</td>
</tr>
<tr>
<td>Specificity</td>
<td>51</td>
<td>77</td>
<td>67</td>
</tr>
<tr>
<td>F1-Score</td>
<td>43</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>AUC</td>
<td>55</td>
<td>59</td>
<td>58</td>
</tr>
<tr>
<td>AUCPR</td>
<td>57</td>
<td>42</td>
<td>43</td>
</tr>
</tbody>
</table>

**Abbreviations:** AUC, area under receiver operating characteristic curve; AUPRC, area under the precision-recall curve (AUPRC). The reason behind lower specificity ratio is due to sensitivity-based training adopted to be relevant in clinical operations; a lower specificity (minimum of 50%) is accepted to reach higher sensitivity.

### 3.3. Prototype Design

The findings of the usability tests with clinicians resulted in an interface divided into four parts (Figure 2b). We identified a need to adapt the model score to the right context to be understandable. Presenting the binary output (e.g. high risk vs low risk) together with the model score resulted in some misinterpretations, although both might still be needed for the information to be understandable. In one design iteration we also included information on the model confidence (i.e. how often the model was correct on the training data for a given model score interval), but this was perceived as too complex by the users. Further, we noticed that additional information was needed for the users to trust and use the SHAP values, e.g. previous lab values and admission reasons. We also recognized that parameters that were converted into multiple binary features, such as normal and abnormal, would be more understandable if grouped together in the user interface. Finally, the language we use and what information we choose to present are important for the user’s overall understanding of and trust in the result.

![Figure 2](image)

Figure 2. From (a) model prediction of a correct readmission case with 74% probability and list of important features from SHAP to (b) a low-fidelity sketch of the user interface. (a) Color indicates which features increase (red) or decrease (blue) the readmission risk, value on x-axis reflects strength of impact on prediction, features are sorted in y-axis according to their importance. (b) 1) Main information: Risk assessment and potentially in-depth explanation of meaning of model output and confidence. 2) SHAP values: Explanation followed by features with top 5 pos/neg SHAP values. Possible to access information about other contributing features. 3) Details about each feature (history and definitions). 4) Access to general model information and possibly links to additional information about model and results.

### 4. Discussion and Conclusion

This paper brings up a use case of working with human-centered AI for healthcare. This type of systems (i.e. CDS applications) are intended to improve healthcare delivery by enhancing medical decisions with knowledge extracted from patient data, targeted clinical knowledge, and other health information. Development of such tools requires the skills of multiple experts across different domains, i.e., medical, artificial intelligence,
implementation science, and UX design, however, also requires cross-disciplinary understanding to translate research outputs into new information [5].

The development process, although agile and adaptive, might face some challenges due to uncertainty caused by the complexity of the healthcare context [12]. One such example was related to what point in the care process to place the CDS application, an aspect that needs definition already in the model development process. Project progress showed several alternatives, due to the organisation of the care process and specific context, requiring further iterations with stakeholders, in combination with objective evaluation of the AI-model performance. This, however, is still in silico, as we cannot clinically test a solution at its infant stages.

Other challenges were related to model output presentation, such that the tool and the XAI can be understood and provide the value sought after by users, as well as continuous dialogues with clinical professionals. Expert knowledge integration into the project was a precondition to develop and design both the AI-model and the interface prototype, thus required careful planning to make the most out of the limited time together with clinical professionals.

In conclusion, the project showed that developing CDS applications are feasible and embraced by potential users. However, there are uncertainties and challenges one should carefully consider and plan for in this type of project.

Acknowledgment and Ethical considerations
This project is part of the CAISR Health research profile funded by the Knowledge foundation. The data analysis is approved by the Swedish Ethics Review Authority (Etikprövningsmyndigheten) 2016/517 with Dr. Markus Lingman as chairperson.

References


