Using Automated Feature Selection for Building Case-Based Reasoning Systems: An Example from Patient-Reported Outcome Measurements

Deepika Verma[1], Kerstin Bach[1], and Paul Jarle Mork[2]

[1] Department of Computer Science
[2] Department of Public Health and Nursing
Norwegian University of Science and Technology, Trondheim, Norway
{deepika.verma, kerstin.bach, paul.mork}@ntnu.no

Abstract. Feature selection for case representation is an essential phase of Case-Based Reasoning (CBR) system development. To (semi-)automate the feature selection process can ease the knowledge engineering process. This paper explores the feature importance provided for XGBoost models as basis for creating CBR systems. We use Patient-Reported Outcome Measurements (PROMs) on low back pain from the SELFBACK project in our experiments. PROMs are a valuable source of information that capture physical, emotional as well as social aspects of well-being from the perspective of the patients. Leveraging the analytical capabilities of machine learning methods and data science techniques for exploiting PROMs have the potential of improving decision making. This paper presents a two-fold approach employed on our dataset for feature selection that combines statistical strength with data-driven knowledge modelling in CBR and compares it with permutation feature selection using XGBoost regressor. Furthermore, we compare the performance of the CBR models, built with the selected features, with two machine learning algorithms for predicting different PROMs.

Keywords: Case-Based Reasoning, Feature Selection, Case Representation, Patient-Reported Outcome Measurements

1 Introduction

Patient-reported outcome measurements (PROMs)[3] are collected routinely in clinical settings and are designed to capture the patients' perception of their own health through structured questionnaires. By utilising machine learning methods and data science techniques, there is a large potential for PROMs to inform and improve clinical decision making [27]. In the current work, we use PROMs on low back pain (LBP) as an example. Among patients seen in primary care, a specific cause of LBP can rarely be identified and the symptoms are most often diagnosed as being "nonspecific". This also highlights the multi-factorial

[3] https://www.hss.edu/proms.asp

nature of LBP, i.e., both genetic, physiological, social and psychological factors are likely to contribute to LBP. While an early and thorough assessment of LBP is recommended (for example, to detect cases at high risk of poor outcome) [18], there are currently no clinical decision support systems (CDSS) in use in clinical practice that can assist or improve such detection or predict the likely outcome for a patient.

Case-Based Reasoning (CBR) systems are well suited for the task of CDSS [5] since the PROMs of the patients can be described in a case-base, a knowledge repository that can aid decision making [2]. However, clinical datasets with PROMs usually contain several clinical measures, all of which may not necessarily be required for decision making and it is therefore necessary to be able to select optimal subset of features that can be used for building CBR systems to predict the patient outcomes and facilitate decision making [10].

Retrieval of similar cases is an important phase in CBR systems, which relies on the case representation and similarity measures. Hence, the selection of the most relevant and important features can easen and simplify the development of the entire CBR system. The focus of this paper is the feature selection phase for building CBR systems from PROMs to predict patient outcomes. While the overall method can be applied to other domains, we will present our evaluation using a dataset with PROMs (described in section 3) in this work. We employ a two-fold approach on our dataset for feature selection that combines statistical strength with data-driven knowledge modelling in CBR and compare it with permutation feature selection using XGBoost regressor. Additionally, we compare the performance of the CBR models, built with the selected features, with two machine learning algorithms for predicting different PROMs.

2 Related Work

PROMs are a valuable source of information and present opportunities for highly sophisticated analysis, but has only been exploited by a few studies in the context of leveraging analytical capabilities of machine learning methods. Rahman et al. [20] used a total of 130 PROMs collected via their pain self-management mobile application ("Manage My Pain"). Using Random Forest, they showed that pain volatility levels at 6 months follow-up could be predicted with a 70% accuracy. In their followup work [19], the authors showed that similar level of accuracy (68%) could be obtained with just 9 features. In another study, Harris et al. [12] used preoperative PROMs to predict whether or not a patient achieves a clinically important improvement in several pain- and function-related outcomes at 1-year post knee arthroplasty. Using several supervised machine learning algorithms, they showed that similar performance can be achieved across different algorithms for the outcomes by varying the number of inputs.

Using the CBR methodology for clinical datasets has already proven useful in decision making [13]. For building robust decision support CBR systems, sufficient description of the problem is necessary. Knowledge about the importance of various features in the dataset plays an important role in problem descrip-

tion for building CBR systems [1]. Xiong and Funk [28] proposed an approach wherein they assessed the feature subset selection based on the performance of CBR models. Later on, the authors proposed a hierarchical approach to select feature subsets for similarity models [29]. They used individual cases to optimise the possibility distributions in the case base and features were selected based on the magnitude of their parameters in the similarity models. Similar to the feature-selection approach proposed by Li et al. [17], we identify optimal feature subsets for our CBR system by iteratively building CBR systems with different feature subsets and evaluating the performance based on the predictions. While Li et al. used mutual information as a preset criterion for selecting feature subsets and evaluating the subsequent CBR systems, we used correlation. In their previous work [16], Li et al. combined feature reduction using rough set with case selection for handling large datasets. Similarly, Zhu et al. [30] selected reduced feature sets through neighborhood rough set algorithm, a method that has been used widely for feature and case selection in CBR [21,22].

3 selfBACK Dataset

The dataset consists of PROMs collected during the randomised controlled trial (RCT)[4] that tested the effectiveness of the selfBACK [5] DSS [23].

Care-seeking patients in primary care with non-specific LBP were recruited to the study. Patients were screened for eligibility based on a set of criteria. The eligible patients were invited to participate in the RCT and those who accepted the invite answered a baseline questionnaire. The participating patients were randomized into either intervention group or control group. The intervention group had access to the selfBACK DSS mobile application and received tailored self-management plans weekly whereas the control group did not. The participants answered questionnaires at different time-points: (1) (only intervention group) at the end of every week: Tailoring questionnaire, and (2) at the end of 6-weeks, 3-months, 6-months and 9-months: Follow-up questionnaire. The questionnaires consist of measures of *pain intensity, pain self-efficacy, physical activity, functional ability, work-ability, sleep quality, fear avoidance* and *mood*. Additionally, the baseline questionnaire included patient sociodemographics (education, employment and family). Table 1 summarises the information collected from the participants at various time-points. We use the Baseline, Follow-up 1 (after 6 weeks) and Follow-up 2 (after 3-months) PROMs in our evaluation. A detailed account of data collection for the RCT can be found in Sandal et al. [23].

From the dataset, six outcomes were selected as target outcomes: Roland Morris Disability Questionnaire (RMDQ, range: [0,24]), Numeric Pain Rating Scale (NPRS, range: [0,10]), Work-ability index (WAI, range: [0,10]), Pain Self Efficacy Questionnaire (PSEQ, range: [0,60]), Fear Avoidance Belief Questionnaire (FABQ, range: [0,30]) and Global Perceived Effect Scale (GPE, range:

[4] `https://clinicaltrials.gov/ct2/show/NCT03798288`
[5] `http://www.selfback.eu`

Table 1: The selfBACK dataset created consists of participant characteristics collected at different time points and includes a selection of PROMs.

Descriptive variables		
Patient Characteristics	Sociodemographics	
Primary Outcome Measure		
Roland Morris Disability Questionnaire		
Secondary Outcome Measures		
Pain Self-Efficacy Questionnaire	Fear Avoidance Belief Questionnaire	Pain Intensity
Brief Illness Perception Questionnaire	Saltin-Grimby Physical Activity Level Scale	
Global Perceived Effect		
Other Outcome Measures		
Workability	Health-related Quality of Life	Activity Limitation
Patient Health Questionnaire	Perceived Stress Scale	Sleep
Patient Specific Functional Scale	Pain Duration and frequency	Physical Activity
Exercise		

[-5,+5]). The primary outcome, RMDQ, is used to evaluate the effect of the self-management app in the RCT. The other outcomes were chosen to elucidate the variation in LBP symptoms amongst the participants.

The intervention group dataset consists of PROMs from 218 participants while the control group dataset contains PROMs of 158 participants. Each participant is initially described by 47 features. Only the participants who completed at least the first two follow-up questionnaires were included in this work.

4 Feature Engineering for CBR systems

Feature selection is an important step in the process of developing CBR systems. Reducing the dimensionality of the data enables the algorithm(s) to train faster by removing redundant information, thereby reducing model complexity, risk of overfitting, better generalisation and aiding interpretability of the models [7]. This is especially pertinent for building CBR systems for datasets with a high dimensionality, such as healthcare-oriented datasets, to ensure focus on the relevant attributes and enhance explainability of the models. Nonetheless, the methodology we present can be used for other domains for feature selection since the principle here is determining the best representation of a dataset in order to learn a solution to a given problem. While we use a healthcare domain dataset, the methodology itself has a broader application.

We use both filter and embedded methods in this work to determine reduced sets of predictors for the target outcomes. *Filter methods* use the principal criteria of ranking technique to select the most relevant features. Features are ranked based on statistical scores, correlation in our case, to determine the features' correlation with the outcome variable. This method is computationally efficient and does not rely on learning algorithms which can introduce a biased feature

subset due to over-fitting [7]. However, correlation-based feature selection has shortcomings if there is a high degree of mutual correlation in the feature set. *Embedded methods* on the other hand are algorithm-specific, iteratively extracting features which contribute the most to the training of a particular iteration of a model during the training process. Impurity-based feature selection using tree-based algorithms[6] is a commonly used embedded method. Permutation feature importance determines the influence of random permutation of each predictor's values on the model performance while still preserving the distribution of the feature [9].

We experimented with two methodologies for selecting optimal predictors for each target outcome:

1. **Correlation and CBR**: Using a two-step hybrid method that combines statistical strength with data-driven case modelling, we attempted to derive optimal predictors of the target outcomes by computing correlation and iteratively building CBR models using features derived from correlation. Here, similarity measure development and building case representation are important factors in evaluating the performance of the CBR models for each set of features.

2. **Permutation feature importance using XGBoost**: Features are selected by computing permutation feature importance (PFI) with XGBoost (XGB) algorithm based on an evaluation metric.

Both methodologies aim to select optimal feature sets based on the trade-off between model performance and model simplicity, that is, fewer features.

4.1 Feature Selection and CBR System Optimization

To determine the optimal set of predictors for developing CBR systems, we experimented with two methodologies for selecting features: correlation-based and based on the feature importance of a XGBoost model. The features selected by both methodologies were used to build CBR systems for all the outcomes at both follow-up time-points. Additionally, we implemented Support Vector and XGB Regression models to compare and contrast the performance of the CBR systems. Figure 1 illustrates the process of feature selection methods we used.

The modeling of the CBR systems was done with the myCBR workbench [3]. The experiments were run using myCBR Rest API[7] [4] for batch querying the CBR systems and python packages such as Scikit learn [6] and XGBoost [8] (python version 3.6.7) were used for building regression models and Pingouin for the statistical correlation [24]. For each target outcome we created datasets with the baseline data as input features and the PROMs of follow-up 1 and follow-up

[6] `https://scikit-learn.org/stable/auto_examples/ensemble/plot_forest_importances.html`

[7] `https://github.com/ntnu-ai-lab/mycbr-sdk`

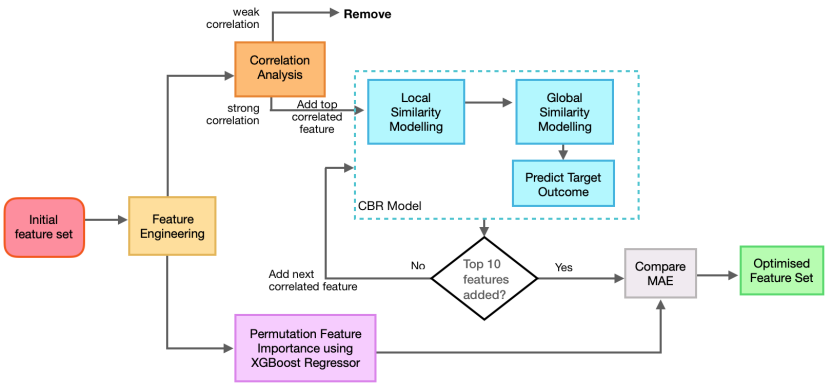


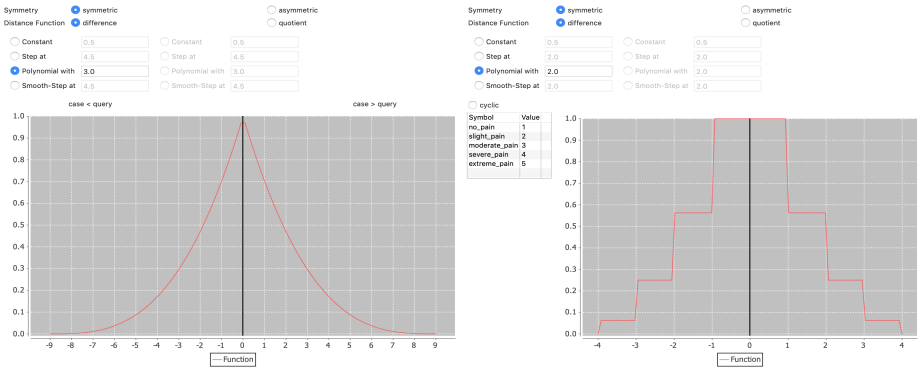
Fig. 1: Flowchart of the feature selection process

2 as target values. These datasets were used to build CBR systems in a data-driven manner and as training data in the other two regression algorithms. In all the CBR models built for various target outcomes in this work, local similarity modelling of the attributes has been done in the same data-driven manner as presented in our previous work [25,26]. The individual features are weighted equally in the global similarity function. Figure 2 showcases examples of local similarity measure modelling for numerical and categorical (ordinal) attributes (using correlated features of NPRS at follow-up 2 as an example). We urge the reader to refer to the previous work to fully grasp how the local similarity measures have been developed, as it is not possible to include the details in this work. Figure 3 shows the case representation of the same target outcome (NPRS) in myCBR workbench with 10 most correlated features.

To predict the target outcomes for a given participant using CBR model, we exploit the *"similar problems have similar solutions"* principle of CBR. While the query participant has been left out (leave-one-out cross validation), we determine their *n-nearest neighbours* (most similar case) with n in range [1,20] and compute mean of the target value reported by the n-neighbours, which is assigned as prediction for the given participant. The process is repeated for each participant and each target outcome dataset at both follow-up time-points for both the intervention and control group. The mean absolute error (MAE) is used as the metric to evaluate the predictive performance of the models.

4.2 Correlation-based Feature Selection

Figure 1 shows that we we first compute correlation between the baseline features and each target outcome to select features. Since the dataset comprises of both numerical and categorical features, we use Pearson for numerical features and one-way ANOVA for categorical features to determine correlation between the baseline features and the target outcomes. Features with absolute correlation greater than the average correlation of the feature set and $p < 0.05$ were



(a) Baseline NPRS (b) EQ5D Pain

Fig. 2: Modelling of Local similarity measures for numerical (a) and categorical (b) attributes in myCBR workbench.

Instance information

Name	Patient1

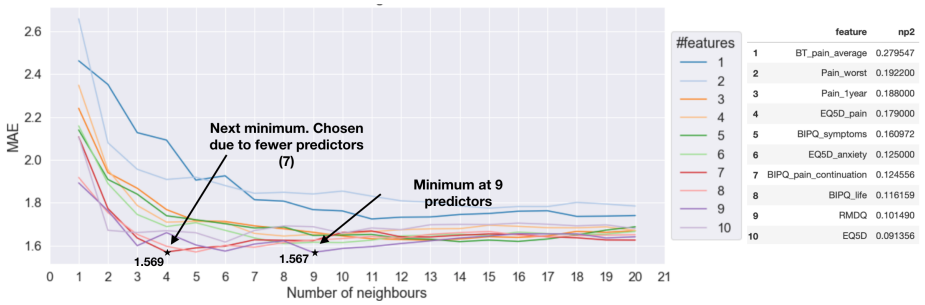
Attributes

BIPQ_life	6.0	Special Value: none
BIPQ_pain_continuation	10.0	Special Value: none
BIPQ_symptoms	5.0	Special Value: none
BT_pain_average	3.0	Special Value: none
EQ5D	80.0	Special Value: none
EQ5D_anxiety	not_anxious	Change Special Value: none
EQ5D_pain	slight_pain	Change Special Value: none
Pain_1year	Above30days	Change Special Value: none
Pain_worst	7.0	Special Value: none
RMDQ	7.0	Special Value: none

Fig. 3: Case representation in myCBR for NPRS (at follow-up 2, control group dataset) with 10 most correlated features

selected. For several reasons including simplified process of modelling in myCBR and based on experience from earlier experiments, it was decided to include only the top ten correlated features for building CBR systems. Previous experiments on the intervention group datasets showed that no more than ten features are necessary to predict any of the chosen target outcomes without any loss in the model performance. To build each CBR model, the casebase is populated with cases imported from a csv file in the myCBR workbench. Local similarity measures are developed for each attribute individually. Instead of building a new CBR model for each set of features, we build one model with the ten most correlated features and use ten different global similarity functions to progressively add more features. Once both the local and the global similarity measures are in place, we batch query the casebase using POST calls in the python implemen-

tation to generate predictions for the target outcome. The MAE is calculated between the reported outcome and the predictions for the entire dataset.



	feature	np2
1	BT_pain_average	0.279547
2	Pain_worst	0.192200
3	Pain_1year	0.188000
4	EQ5D_pain	0.179000
5	BIPQ_symptoms	0.160972
6	EQ5D_anxiety	0.125000
7	BIPQ_pain_continuation	0.124556
8	BIPQ_life	0.116159
9	RMDQ	0.101490
10	EQ5D	0.091356

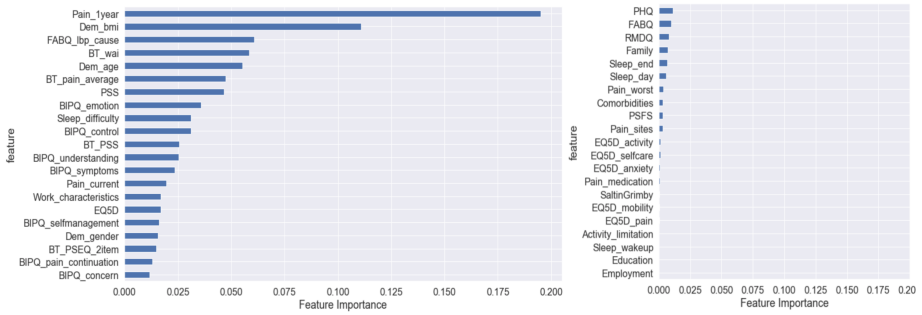
Fig. 4: On the right side of the figure are the top ten correlated features used to build the CBR model for predicting *NPRS* (input: baseline data, target: NPRS at follow-up 2). Features were added progressively one at a time in the given order, starting with the most correlated feature. **np2** (eta-squared) is the squared correlation coefficient. Graph on the left shows the MAE variation with different sets of features in the corresponding CBR model for predicting *NPRS*, with x-axis presenting the n-neighbours used for generating predictions and y-axis presenting the MAE in the predictions for the entire dataset.

Figure 4 gives an example for one target outcome, *NPRS*. It shows the result of the correlation (left) and the MAE when predicting the NPRS using the baseline data (right). We can see that the progressive addition of correlated features improves the prediction by the CBR system already by using the most similar case. Further, we observe that adding neighbors generally reduces the error and for the final model we choose the combination with the lowest MAE.

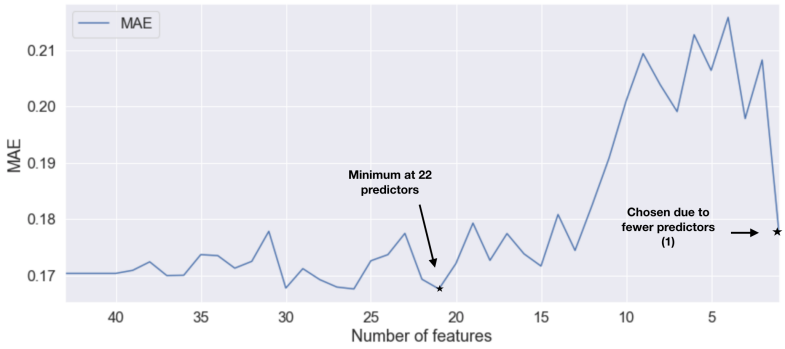
4.3 Feature Importance using XGBoost

In this approach, we select features by computing the permutation feature importance using the XGBoost Regressor and compare the MAE of the predictions to determine the optimal feature set. The permutation feature importance is determined by the difference between the modified (permuted) dataset and a baseline model based on the MAE. First, a baseline model with all the features is trained and its MAE is computed. Next, the values of one feature in the dataset are permuted and then the model is re-trained and the MAE is computed for the modified dataset. The process is repeated for all the features in the dataset. The optimal number of features are selected based on the trade-off between model performance and number of features.

Figure 5a shows the feature importance for predicting the GPE and figure 5b shows the development of the MAE while adding the features. To select the best configuration, we choose the set with the lowest number of features that has



(a)



(b)

Fig. 5: Feature Selection using permutation feature importance with XGB for predicting *GPE* (input: baseline data, target: GPE at follow-up 1). **a.** Features ranked by their importance. **b.** Effect of feature permutation on the XGB model: The MAE on the y-axis in this plot is scaled.

the lowest MAE as shown in figure 5b. We favor the lowest number of features to build simpler model that requires minimal data collection and can be better explained. The selected features are then used to build CBR model in exactly the same way as described in the previous section and the prediction results are noted.

5 Experimental Results

To compare the performance of the CBR systems, we implemented two regression algorithms, XGB and Support Vector Regression (SVR) for each corresponding CBR system to predict the target outcomes. The algorithms were selected based on previous experiments with the intervention group data where we evaluated the performance of XGB and SVR along with other algorithms, including Linear Regression, Passive Aggressive Regression, Stochastic Gradient Descent, Ad-

aBoost, Random Forest, and found SVR and XGB to lead to the best results. For the simplicity of comparison and clarity, it was decided to keep only SVR and XGB for further evaluation. To optimize the hyperparameters, we used grid search [15]. Tables 2 and 3 summarise the results of predicting target outcomes using the CBR models, SVR and XGB for the intervention and control group participants, respectively.

Table 2: Results of Prediction of Target Outcomes using different Feature Selection Methodologies and Regression Methods for the Intervention Group (size of dataset: 218 participants). Numbers in bold letters are lowest MAE. **FU1**: Follow-up 1, **FU2**: Follow-up 2, **n**: number of features

		Feature Selection Methodology							
		Correlation+CBR				**PFI+XGBoost**			
Target	**Follow-Up**	**n**	**CBR**	**SVR**	**XGB**	**n**	**CBR**	**SVR**	**XGB**
RMDQ	**FU1**	4	2.98	3.19	3.32	5	2.78	**2.69**	2.71
	FU2	8	2.90	**2.83**	2.85	4	3.17	3.92	3.02
NPRS	**FU1**	7	**1.38**	1.45	1.50	3	1.50	1.49	1.52
	FU2	9	1.48	**1.33**	1.38	3	1.46	1.41	1.42
WAI	**FU1**	5	1.16	1.98	1.98	2	**1.14**	1.96	2.01
	FU2	4	**1.14**	2.16	2.21	1	1.24	2.19	2.24
PSEQ	**FU1**	1	5.50	16.9	17.0	2	**5.45**	17.2	17.3
	FU2	3	5.95	16.6	16.6	2	**5.95**	16.4	17.1
FABQ	**FU1**	3	3.87	3.74	3.76	6	3.90	**3.50**	3.67
	FU2	1	3.9	**3.60**	3.84	6	3.83	3.64	3.86
GPE	**FU1**	1	**1.37**	2.73	2.76	2	1.39	2.82	2.78
	FU2	2	1.54	2.51	2.43	3	**1.49**	2.54	2.46

6 Discussion

A number of inferences can be made based on the results. We see in figure 4 that the baseline measurement (listed as *BT_pain_average*) of the associated target outcome *NPRS* is its' first most important predictor. This is a trend observed for all the target outcomes, except *GPE* which does not have an associated baseline measurement (see figure 5b). This is an important observation from clinical perspective, since baseline measurements of the associated outcomes have previously been found to be their most important predictor [11,14], and our experiments support these findings.

Table 3: Results of Prediction of Target Outcomes using different Feature Selection Methodologies and Regression Methods for the Control Group (size of dataset: 158 participants). Numbers in bold letters are lowest MAE. **FU1**: Follow-up 1, **FU2**: Follow-up 2, **n**: number of features

		Feature Selection Methodology							
		Correlation+CBR				**PFI+XGBoost**			
Target	**Follow-Up**	**n**	**CBR**	**SVR**	**XGB**	**n**	**CBR**	**SVR**	**XGB**
RMDQ	**FU1**	2	3.11	2.99	2.97	4	3.07	2.92	**2.75**
	FU2	2	3.11	**2.97**	3.14	3	3.22	2.97	3.14
NPRS	**FU1**	6	**1.41**	1.77	1.85	2	1.49	1.73	1.85
	FU2	7	1.56	**1.49**	1.7	1	1.72	1.56	1.71
WAI	**FU1**	1	1.02	1.02	**1.01**	1	1.02	1.02	1.01
	FU2	2	1.14	**1.12**	1.17	1	1.19	1.15	1.18
PSEQ	**FU1**	7	**6.68**	19.2	19.6	1	7.01	19.4	19.8
	FU2	3	6.23	19.0	19.5	5	**5.94**	19.1	19.3
FABQ	**FU1**	1	3.47	**3.27**	3.58	1	3.47	3.27	3.58
	FU2	2	3.77	3.69	3.80	2	3.82	**3.58**	3.93
GPE	**FU1**	7	**1.22**	2.55	2.52	1	1.26	2.61	2.49
	FU2	1	**1.33**	2.65	2.58	2	1.39	2.67	2.56

Selecting optimal features, especially for healthcare datasets, is one of those application domains where no one particular method prevails and one must decide based on application domain knowledge and experience, among others. From the results in table 2 and 3, we see that the features selected by either of the methodologies give similar results with respect to the error in predictions. There is no clear winner here. However, taking into consideration the time and effort required, XGBoost permutation feature importance methodology requires minima and provides a more streamlined process for selecting optimal feature sets as compared to the two-fold approach, which requires estimating correlation, building several similarity measures and CBR systems for the target outcomes and comparing the MAE for determining optimal feature sets. As for a concrete time comparison, it is not possible since the modelling of local and global similarity measures for building a CBR model requires manual input. On the other hand, this comparison also establishes the utility of the two-fold approach for building tailored CBR systems.

All the three regression methods give fairly similar results when it comes to predicting the outcomes. However, for an outcome with a relatively large range (*PSEQ*) or no baseline measurement of the target outcome (*GPE*), both SVR and XGB fall short in comparison to the results we get from the CBR models. This is similar to our findings in our previous work [25] where we found

CBR model built with our data-driven modelling approach to be more sensitive and robust to the data-distribution of individual features, thereby, furthering our premise that both data-driven similarity modelling and CBR are better suited for this task. Moreover, outcomes generated by CBR models are more explainable, which is a pre-requite for any CDSS where explainable systems are preferred over complex ones.

7 Conclusion and Future Work

In this paper, we presented a two-fold approach for feature selection wherein we used the correlation coefficient as a pre-processing step to select ten most correlated features and build the CBR models with progressively more features for predicting PROMs. We examine the performance of the predictions generated using CBR systems to determine optimal feature subsets for the outcomes. Through evaluation and comparison with tree-based feature selection methods (permutation feature importance with XGBoost), it can be concluded that although the presented two-fold approach is feasible and gives results similar to the other approach undertaken, it is however more time and effort intensive and therefore, feature selection using XGBoost permutation feature importance appears to be a more promising option. Predictive performance of the CBR systems is at par with and many a times better than the traditional algorithms such as SVR and XGBoost.

From a clinical perspective, building prognostic models that can provide necessary information to clinicians and patients of possible outcome(s) pertaining to a specific treatment is a necessity to support informed clinical decision making. Access to individualized predictive analytics for different outcomes may be the next step in the management of pain and related symptoms for patients with LBP. The results we get from our dataset confirm the predictive value of baseline measurements of associated target outcomes, similar to other studies such as by Fontana et al. [11] and Huber et al. [14].

In future work, it may be worthwhile to compare performance of the CBR models built with features selected by an expert with the approach presented in this work.

Acknowledgement The work has been conducted as part of the SELFBACK and Back-UP projects, which have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 689043 and No 777090.

References

1. Aamodt, A., Plaza, E.: Case-based reasoning: Foundational issues, methodological variations, and system approaches. Artificial Intelligence Communications, 7(1), 39–59 (1994)

2. Andritsos, P., Jurisica, I., Glasgow, J.I.: Case-based reasoning for biomedical informatics and medicine. In: Springer Handbook of Bio-/Neuroinformatics, pp. 207–221. Springer (2014)
3. Bach, K., Althoff, K.D.: Developing case-based reasoning applications using mycbr 3. In: International Conference on Case-Based Reasoning. pp. 17–31. Springer (2012)
4. Bach, K., Mathisen, B.M., Jaiswal, A.: Demonstrating the mycbr rest api. In: ICCBR Workshops. pp. 144–155 (2019)
5. Bichindaritz, I., Marling, C.: Case-Based Reasoning in the Health Sciences: Foundations and Research Directions, vol. 309, pp. 127–157. Springer (05 2010)
6. Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., Varoquaux, G.: API design for machine learning software: experiences from the scikit-learn project. In: ECML PKDD Workshop: Languages for Data Mining and Machine Learning. pp. 108–122 (2013)
7. Chandrashekar, G., Sahin, F.: A survey on feature selection methods. Computers & Electrical Engineering 40(1), 16–28 (2014)
8. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. p. 785–794. KDD '16, Association for Computing Machinery, New York, NY, USA (2016)
9. Fisher, A., Rudin, C., Dominici, F.: All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. Journal of Machine Learning Research 20(177), 1–81 (2019)
10. Floyd, M.W., Davoust, A., Esfandiari, B.: Considerations for real-time spatially-aware case-based reasoning: A case study in robotic soccer imitation. In: European Conference on Case-Based Reasoning. pp. 195–209. Springer (2008)
11. Fontana, M.A., Lyman, S., Sarker, G.K., Padgett, D.E., MacLean, C.H.: Can machine learning algorithms predict which patients will achieve minimally clinically important differences from total joint arthroplasty? Clinical Orthopaedics and Related Research 477(6), 1267–1279 (2019)
12. Harris, A.H., Kuo, A.C., Weng, Y., Trickey, A.W., Bowe, T., Giori, N.J.: Can machine learning methods produce accurate and easy-to-use prediction models of 30-day complications and mortality after knee or hip arthroplasty? Clinical orthopaedics and related research 477(2), 452 (2019)
13. Holt, A., Bichindaritz, I., Schmidt, R., Perner, P.: Medical applications in case-based reasoning. Knowledge Engineering Review 20(3), 289–292 (2005)
14. Huber, M., Kurz, C., Leidl, R.: Predicting patient-reported outcomes following hip and knee replacement surgery using supervised machine learning. BMC medical informatics and decision making 19(1), 3 (2019)
15. Hutter, F., Kotthoff, L., Vanschoren, J.: Automated Machine Learning. Springer (2019)
16. Li, Y., Shiu, S.C.K., Pal, S.K., Liu, J.N.K.: A rough set-based case-based reasoner for text categorization. International journal of approximate reasoning 41(2), 229–255 (2006)
17. Li, Y.F., Xie, M., Goh, T.: A study of mutual information based feature selection for case based reasoning in software cost estimation. Expert Systems with Applications 36(3), 5921–5931 (2009)
18. Lin, I., Wiles, L., Waller, R., Goucke, R., Nagree, Y., Gibberd, M., Straker, L., Maher, C.G., O'Sullivan, P.P.: What does best practice care for musculoskeletal

pain look like? eleven consistent recommendations from high-quality clinical practice guidelines: systematic review. British journal of sports medicine 54(2), 79–86 (2020)
19. Rahman, Q.A., Janmohamed, T., Clarke, H., Ritvo, P., Heffernan, J., Katz, J.: Interpretability and class imbalance in prediction models for pain volatility in manage my pain app users: analysis using feature selection and majority voting methods. JMIR medical informatics 7(4), e15601 (2019)
20. Rahman, Q.A., Janmohamed, T., Pirbaglou, M., Clarke, H., Ritvo, P., Heffernan, J.M., Katz, J.: Defining and predicting pain volatility in users of the manage my pain app: Analysis using data mining and machine learning methods. Journal of medical Internet research 20(11), e12001 (2018)
21. Salamó, M., Golobardes, E.: Rough sets reduction techniques for case-based reasoning. In: International Conference on Case-Based Reasoning. pp. 467–482. Springer (2001)
22. Salamo, M., Lopez-Sanchez, M.: Rough set based approaches to feature selection for case-based reasoning classifiers. Pattern Recognition Letters 32(2), 280–292 (2011)
23. Sandal, L.F., Stochkendahl, M.J., Svendsen, M.J., Wood, K., Øverås, C.K., Nordstoga, A.L., Villumsen, M., Rasmussen, C.D.N., Nicholl, B., Cooper, K., Kjaer, P., Mair, F.S., Sjøgaard, G., Nilsen, T.I.L., Hartvigsen, J., Bach, K., Mork, P.J., Søgaard, K.: An app-delivered self-management program for people with low back pain: Protocol for the selfback randomized controlled trial. JMIR Res Protoc 8(12), e14720 (Dec 2019)
24. Vallat, R.: Pingouin: statistics in python. Journal of Open Source Software 3(31), 1026 (2018), https://doi.org/10.21105/joss.01026
25. Verma, D., Bach, K., Mork, P.J.: Modelling similarity for comparing physical activity profiles - a data-driven approach. In: Cox, M.T., Funk, P., Begum, S. (eds.) CBR Research and Development. Springer, Cham (2018)
26. Verma, D., Bach, K., Mork, P.J.: Similarity measure development for case-based reasoning–a data-driven approach. In: Bach, K., Ruocco, M. (eds.) Nordic Artificial Intelligence Research and Development. pp. 143–148. Springer International Publishing, Cham (2019)
27. Wu, A., Kharrazi, H., Boulware, L., Snyder, C.: Measure once, cut twice -adding patient-reported outcome measures to the electronic health record for comparative effectiveness research. Journal of clinical epidemiology 66, S12–20 (08 2013)
28. Xiong, N., Funk, P.: Construction of fuzzy knowledge bases incorporating feature selection. Soft Computing 10(9), 796–804 (2006)
29. Xiong, N., Funk, P.: Combined feature selection and similarity modelling in case-based reasoning using hierarchical memetic algorithm. In: IEEE Congress on Evolutionary Computation. pp. 1–6. IEEE (2010)
30. Zhu, G.N., Hu, J., Qi, J., Ma, J., Peng, Y.H.: An integrated feature selection and cluster analysis techniques for case-based reasoning. Engineering Applications of Artificial Intelligence 39, 14–22 (2015)