informatics

MDPI

Review

Application of Machine Learning Methods on Patient Reported Outcome Measurements for Predicting Outcomes: A Literature Review

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

[1] Department of Computer Science, Norwegian University of Science and Technology, 7034 Trondheim, Norway; kerstin.bach@ntnu.no
[2] Department of Public Health and Nursing, Norwegian University of Science and Technology, 7034 Trondheim, Norway; paul.mork@ntnu.no
* Correspondence: deepika.verma@ntnu.no

Abstract: The field of patient-centred healthcare has, during recent years, adopted machine learning and data science techniques to support clinical decision making and improve patient outcomes. We conduct a literature review with the aim of summarising the existing methodologies that apply machine learning methods on patient-reported outcome measures datasets for predicting clinical outcomes to support further research and development within the field. We identify 15 articles published within the last decade that employ machine learning methods at various stages of exploiting datasets consisting of patient-reported outcome measures for predicting clinical outcomes, presenting promising research and demonstrating the utility of patient-reported outcome measures data for developmental research, personalised treatment and precision medicine with the help of machine learning-based decision-support systems. Furthermore, we identify and discuss the gaps and challenges, such as inconsistency in reporting the results across different articles, use of different evaluation metrics, legal aspects of using the data, and data unavailability, among others, which can potentially be addressed in future studies.

Keywords: machine learning; patient-reported outcome measurements; self-reported measures; patient outcomes; outcome prediction; clinical decision making; decision-support systems; health informatics

Citation: Verma, D.; Bach, K.; Mork, P.J. Application of Machine Learning Methods on Patient Reported Outcome Measurements for Predicting Outcomes: A Literature Review. *Informatics* **2021**, *8*, 56. https://doi.org/10.3390/informatics8030056

Academic Editor: Kamran Sedig

Received: 30 June 2021
Accepted: 19 August 2021
Published: 25 August 2021

1. Introduction

There is growing interest and support for the utility and importance of patient-reported outcome measures (PROMs) in clinical care. PROMs are commonly defined as reports or questionnaires completed by patients to measure their view on their functional well-being and health status [1]. Thus, PROMs may capture the patient's perspective on both social, physical, and mental well-being. Shifting the focus from disease-specific factors towards the patient's perspective may provide a useful basis for shared medical decision making between a clinician and a patient [2,3]. Recent evidence indicates that shared decision making has a positive impact on the quality of decision making, satisfaction with treatment, and patient–provider experience [4]. Likewise, well-informed patients agreeing upon their course of treatment with their caregiver have better outcome and satisfaction [5].

PROMs may play an important role in shared decision making; however, there is currently an unused potential in both collecting and utilising PROMs in clinical practice. Notably, digital innovations can facilitate delivery, storage, processing, and access to PROMs, using third-party or electronic health record (EHR)-based outcome measurement platforms. Intelligent methods can also support shared decision making through digital decision aids and patient engagement platforms, comprising high-quality educational material, and patient–provider communication portals [5,6]. In this context, utilising machine learning and artificial intelligence provide a promising avenue for enhancing the usefulness of PROMs [7].

Several recent studies demonstrated the predictive prowess of machine learning models utilising EHR datasets for the scheduling of surgeries [8–10], and risk stratification [11–13] among others. Singal et al. [14] in their work found the machine learning models to outperform conventional models in predicting the development of hepatocellular carcinoma among cirrhotic patients. The application of machine learning methods on PROMs datasets can allow the exploration of associations in the data that are important for predicting different outcomes, thereby informing a shared decision-making process [15]. Currently, PROMs data are widely used in explanatory research, where researchers typically test hypotheses using a preconceived theoretical construct by applying statistical methods (for example, low back pain is associated to lower quality of life and depression [16,17]. In contrast, PROMs in predictive research can be used to predict outcomes in the future by applying statistical or machine learning methods without any preconceived theoretical constructs (for example, predicting the risk of depression [18]), and is therefore an important step towards patient-centred care with a shift in focus towards the patient's perspective [19].

While prediction models exist that utilise a combination of PROMs and objective clinical data or EHR data for individual predictions [20], models that utilise solely PROMs data to make individual predictions are rare. Despite the broad area of application of machine learning and data science techniques in the biomedical field, the utilisation of these techniques in clinical practice remains low, especially concerning the utilisation of PROMs. A few machine learning applications utilising PROMs data in biomedical research have emerged during recent years; however, the potential for utilising PROMs data to improve clinical care appears under-explored, especially from the perspective of supporting shared decision-making.

The main of aim of this literature review is, therefore, to provide a summary of existing methodologies that apply machine learning methods on PROMs for predicting clinical outcomes and building prognostic models. In Section 2, we introduce the process of article selection and present an analysis of the selected articles in terms of their publication year, intervention domains, length of outcome prediction, data source, feature selection strategy and the machine learning methods used. Furthermore, we discuss the gaps and challenges in Section 3 that can be addressed in future work to utilise machine learning methods on PROMs datasets. The main contribution of this work is firstly, the identification of scientific articles applying machine learning methods on PROMs data for predicting clinical outcomes and secondly, augmenting the utility of machine learning methods for healthcare datasets for building clinical decision support systems to better facilitate decision making for patient-centred care and precision medicine.

2. Methods

2.1. Review Design and Search Strategy

This literature review identifies scientific articles that focus on the application of machine learning methods in the process of predicting short or long-term clinical outcome(s) using PROMs data.

A structured literature search was performed in September 2020, using the following search string in the PubMed and Scopus database: (((self reported measures) OR patient reported measures)) AND ((artificial intelligence) OR machine learning) AND ((outcome prediction) OR outcome assessment). The results were filtered to include journal and conference articles written in English and published within the last decade (2010–2020).

2.2. Article Selection

The following inclusion criteria were used to identify articles relevant for the current review:

- *Data*: The dataset consists of structured questionnaires administered to patients or participants either in-person or via web application before, during and/or after a

treatment. Articles that involved objectively measured data or data gathered from online patient forums were excluded from this study.

- *Machine Learning*: Application of machine learning methods with the intent of data analysis or clustering of patients or assessment of features with prognostic value for one or more target outcomes or building prognostic models for short- or long-term prediction of one or more outcome.
- Full text availability (including institutional access).
- Written in English.

Articles not meeting the inclusion criteria following the abstract and full screening were excluded from this study.

2.3. Search Outcome

Figure 1 presents a flowchart of the article selection process. Based on the structured literature search, a total of 319 records were identified: PubMed (n = 314) and Scopus (n = 5). Further, we screened the references of the articles that met the inclusion criteria along with relevant review articles and books to identify additional articles (n = 4). Finally, after duplicates were removed, we screened 322 articles. After screening of title/abstract and assessing the eligibility, a total of 15 articles were included in the qualitative synthesis.

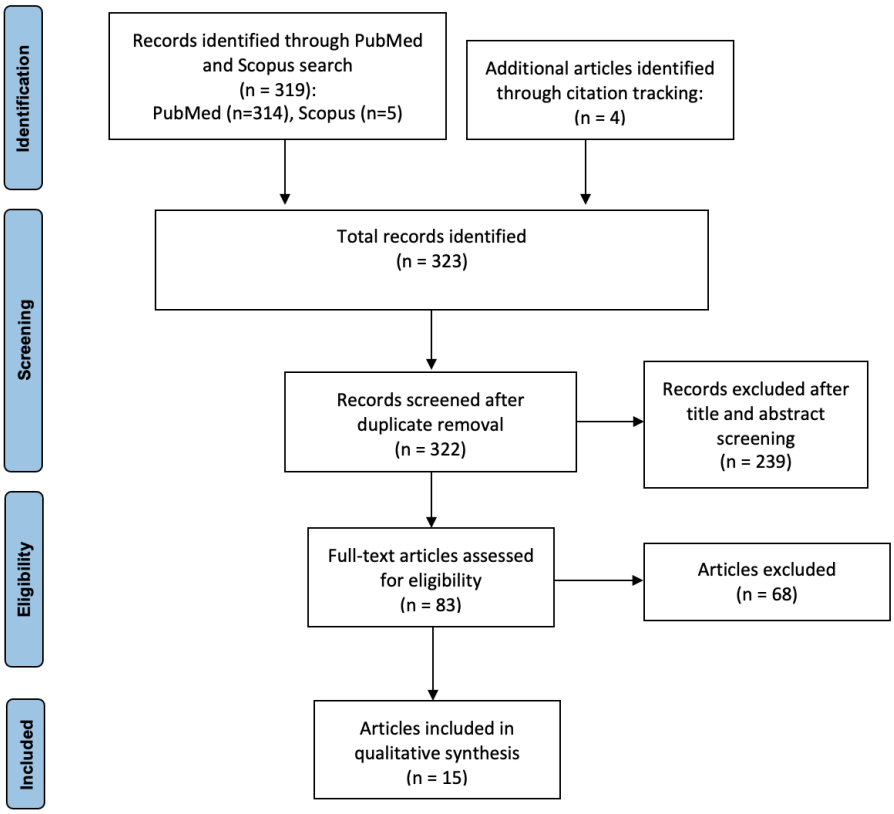

Figure 1. Flowchart of the article selection process.

2.4. Sources of Evidence

All the included articles were published in peer-reviewed journals. A total of 8 out of the 15 articles were published in the years 2019 and 2020 (excluding October–December 2020); see Figure 2. Fourteen articles were published the second half of the decade, 2016–2020, while only one article was published in the first half of the decade, in 2012.

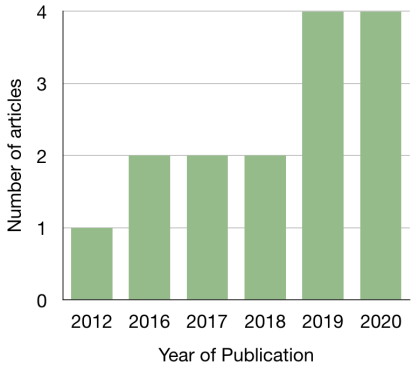

Figure 2. Publication year of included articles.

2.5. Intervention Domains and Length of Prediction

Articles stratified by the intervention domain (Figure 3), can be broadly categorised as post-surgical improvements or limitations, depression, pain management, hospital readmission, and oral health.

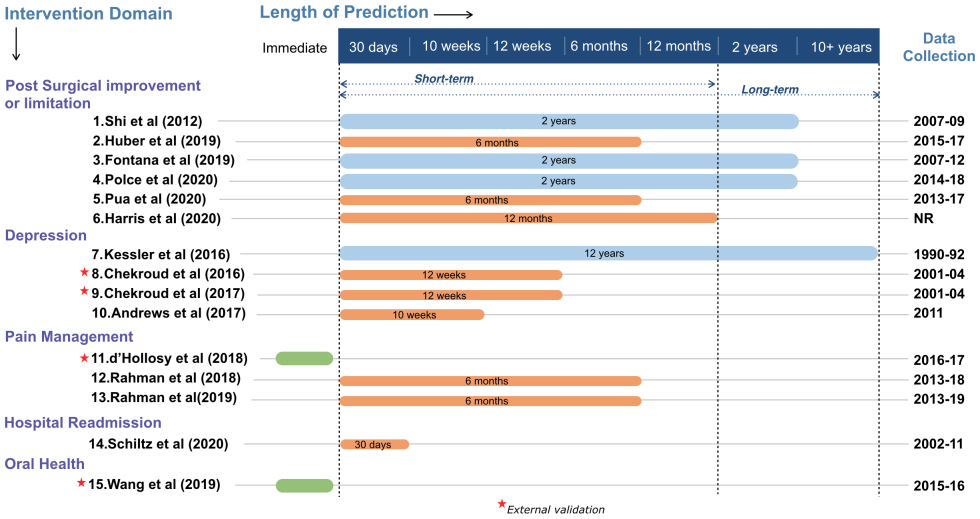

Figure 3. The included articles categorised by their intervention domains. The length of the predictions are indicated, categorised into short- and long-term. The time period of the data collection is indicated to the right. Red asterisks indicate studies that utilised external validation datasets to test the generalisability of the machine learning models.

The first category includes six articles, focusing on outcomes relating to post-surgical limitations or improvements, such as quality of life after cancer surgery [21] and (walking) limitations or improvements (minimal clinically important difference (MCID)) after total joint arthroplasty [22–26]. The second category includes four articles, focusing on identifying patients with depression based on self-reports [18,27] and prognosis of outcome of anti-depression treatment [28,29]. The third category includes three articles focusing on predicting pain volatility amongst users of a pain-management mobile application [30,31] and self-referral decision support for patients with low back pain in primary care [32]. The fourth category includes one article that focused on the risk of hospital readmission [33], while the fifth and last category includes one article that focused on oral health outcome among children aged 2–17 years [34].

Eleven articles presented machine learning models for predicting short-term outcomes (12 months or less), see Figure 3, while four articles presented machine learning models for predicting long-term outcomes (over 12 months). Two articles focused on immediate outcomes, such as referral decision [32] and oral health scores [34]. Four articles, marked with a red asterisk in Figure 3, utilised external validation datasets to test the generalisability of the machine learning models. None of the articles with long-term outcomes utilised external validation datasets. The prediction timelines also appear to be domain dependant. The outcomes from interventions such as depression treatment or surgeries seem to be predicted over the long term, likely due to the nature of the treatment and associated outcomes in the two intervention domains.

2.6. Sources of Data and Availability

Table 1 presents a summary of the included articles. Few articles utilised open-source or available-on-request datasets from national registries, such as National Institute of Mental Health (NIMH) or National Health Service (NHS). The sizes of the datasets vary, from 37 patients [18] to 64,634 patients [22]. Seven articles utilised training datasets with fewer than 1000 patients.

2.7. Feature Selection

The methods of feature selection were either statistical, algorithm-based or manual, based on expertise or availability of data (Table 1). In the table, 'Algorithm implicit' implies that the features were selected by the algorithm(s) used for the prediction task and no other explicit feature selection was carried out, while 'Manual' implies that the features were selected manually based on experience or expert knowledge or data availability.

Ten articles used supervised learning algorithms to extract relevant features from the dataset, while in four articles, features were selected manually, without any statistical or algorithmic assistance. One article [21] applied statistical methods to extract and select relevant features. Among the four articles that employed manual feature selection, two articles [24,34] manually divided all the features into sets and added the sets incrementally into the training dataset to train the model(s). In comparison, in the other two articles [23,26], features were selected manually based on clinical expertise [23] and previous experimental evaluation [26]. Ten articles employed the algorithmic approach for extraction and selection of relevant features from the datasets: Andrews et al. [18] used LASSO; Schiltz et al. [33] and Rahman et al. [31] used Random Forest; Polce et al. [25] used recursive feature elimination with Random Forest; Chekroud et al. [28,29] used Elastic nets; and Huber et al. [22], Rahman et al. [30], d'Hollosy et al. [32] and Kessler et al. [27] employed no separate feature selection but relied on the implicit feature selection ability of the algorithms used. Random Forest and linear models, such as Elastic nets and LASSO, appear to be the preferred algorithm choice for feature selection.

Table 1. Overview of feature selection, model evaluation, data availability and external validation in the included articles. Abbreviations: MCID—minimal clinically important difference, NR—not reported, ANOVA—analysis of variance, CV—cross validation, RoC—receiver operating characteristic, LASSO—least absolute shrinkage and selection operator, NHS—National Health Service, NIMH—National Institute of Mental Health, HRS—Health and Retirement Study.

Article	Outcome	Dataset Size	Total No. of Features	Features Selected	Feature Selection Method	Hyperparameter Tuning	Model Evaluation	Data Availability	External Validation
Shi et al. [21]	Quality of life post surgery	403	NR	NR	ANOVA, Fisher exact analysis, Univariate analysis	NR	Holdout (80,20)	NR	no
Huber et al. [22]	MCID post surgery	64,634	81	NR	Algorithm implicit	NR	5-fold CV	NHS [1]	no
Fontana et al. [24]	MCID post surgery	13,809	NR		Manual	5-fold CV	Holdout (80,20)	NR	no
Polce et al. [25]	Satisfaction post surgery	413	16	10	Recursive Feature Elimination, Random Forest	10-fold CV	Holdout (80,20)	NR	no
Pua et al. [23]	Walking limitation post surgery	4026	NR	25	Manual	5-fold CV	Holdout (70,30)	NR	no
Harris et al. [26]	MCID post surgery	587	NR	NR	Manual	NR	10-fold CV, bootstrapping	NR	no
Kessler et al. [27]	Depressive Disorder chronicity, persistence, severity	1056	NR	9–13	Ensemble and Penalised Regression	NR	10-fold CV	NR	no
Chekroud et al. [28]	Antidepressant treatment	1949	164	25	ElasticNet	RoC maximisation	10*Repeated 10-fold CV	NIMH [2]	yes
Chekroud et al. [29]	Antidepressant treatment	7221	164	25	ElasticNet	NR	5-fold CV	NIMH [2]	yes
Andrews et al. [18]	Depression in older adults	37	6	2	LASSO	Stratified CV	5-fold CV	NR	no
d'Hollosy et al. [32]	Low Back pain self-referral	1288	15	NR	Algorithm implicit	NR	Holdout (70,30)	On Request	yes
Rahman et al. [30]	Pain volatility	782	130	NR	Algorithm implicit	NR	5-fold CV	NR	no

Table 1. *Cont.*

Article	Outcome	Dataset Size	Total No. of Features	Features Selected	Feature Selection Method	Hyperparameter Tuning	Model Evaluation	Data Availability	External Validation
Rahman et al. [31]	Pain volatility	879	132	9	Gini impurity, Information gain, Class imbalance	NR	5-fold CV	NR	no
Schiltz et al. [33]	Hospital Readmission	6617	NR	NR	Random Forest	NR	Holdout (80,20)	HRS [3]	no
Wang et al. [34]	Oral Health	908	27	NA	Manual	Greedy approximation [35]	Holdout (70,30)	NR	yes

[1] https://digital.nhs.uk/data (first accessed on 14 October 2020); [2] https://www.nimhgenetics.org/download-tool/DP (accessed on 9 October 2020); [3] https://hrsonline.isr.umich.edu. (accessed on 16 October 2020).

2.8. Trends in the Application of Machine Learning Methods

Table 2 presents an overview of the different machine learning methods used in the included articles. Ensembles and linear methods appear to be the most commonly applied methods to the PROMs datasets, with all the included articles employing at least either one, likely due to their ability to extract features implicitly. While supervised learning methods are the go-to methods for prediction tasks, three (20%) articles apply unsupervised methods as a pre-step to the supervised methods to determine and predict cluster-specific outcomes [29–31]. Examples of commonly used linear algorithms in the included articles are logistic regression, logistic regression with splines, elastic nets, Poisson regression, LASSO, and linear kernel-based Support Vector Machines, among others. The most commonly applied ensemble algorithms are Random Forest, Boosted Trees, Gradient Boosting Machines (GBM), stochastic gradient boosting machines, extreme gradient boosting (XGBoost), and SuperLearner.

Thirteen (87%) articles used binary classification to predict whether the targeted outcome(s) are above or below a specified threshold (for instance, whether or not a patient achieves MCID in their post-operative outcomes [24]). One article used ternary classification to predict the self-referral outcome among people with low back pain in a primary care setting [32]. In contrast, three (20%) articles used regression [21,29,34], one of which used both regression and binary classification to predict continuous and categorical outcomes [34].

2.9. Study Design and Model Evaluation

To reduce the risk of overfitting the models and to improve their generalisability, a k-fold cross-validation scheme was used in eleven articles, either during the hyperparameter tuning phase or the model evaluation phase (Table 1). Out of these eleven, only one article used the k-fold cross-validation scheme in both phases [18]. Three articles [23,32,34] employed a holdout (70,30) validation approach: 70% of the dataset was used for training the model and 30% for validation, while four articles employed a holdout (80,20) validation approach [21,24,25,33]. While the holdout validation approach is useful due to its speed and simplicity, it often leads to high variability due to the differences in the training and test datasets, which can result in significant differences in the evaluation metric estimates (accuracy, error, sensitivity, etc., depending on the machine learning task the metric used).

External validation datasets were used in four articles to test the generalisability of the models [28,29,32,34]. While external validation is generally recommended to validate the models generated since prediction models perform better on the training data than on new data, internal validation appears to be more common, likely due to either the lack or unavailability of an appropriate external validation dataset. However, to correct the bias in the internally-validated prediction models, bootstrapping methods are recommended [36,37]. Only one article used bootstrapping to internally validate the models where an external validation dataset was not used [26].

Table 2. Overview of the application of different machine learning methods in the included articles. Abbreviations: DT—Decision Tree, SVM—Support Vector Machines, NN—neural network, NB—naive Bayes, k-NN—k-Nearest Neighbour, QDA—Quadratic Discriminant Analysis, Aggl—agglomerative clustering.

Article	Supervised								Unsupervised		Machine Learning Task
	Ensemble Methods	Linear Methods	DT	SVM	NN	NB	k-NN	QDA	k-Means	Aggl.	
Shi et al. [21]		✓			✓						Regression
Huber et al. [22]	✓	✓			✓	✓	✓				Classification
Fontana et al. [24]	✓	✓		✓							Classification
Polce et al. [25]	✓	✓		✓	✓						Classification
Pua et al. [23]	✓	✓									Classification
Harris et al. [26]	✓	✓						✓			Classification
Kessler et al. [27]	✓	✓									Classification
Chekroud et al. [28]	✓										Classification
Chekroud et al. [29]	✓									✓	Regression
Andrews et al. [18]		✓									Classification
d'Hollosy et al. [32]	✓		✓								Classification
Rahman et al. [30]	✓	✓		✓					✓		Classification
Rahman et al. [31]	✓	✓		✓					✓		Classification
Schiltz et al. [33]	✓	✓	✓								Classification
Wang et al. [34]	✓					✓					Regression, Classification

2.10. Model Performance

While it is difficult to provide a concrete result comparison among the included articles due to the utilisation of various metrics, most articles did report at least above chance (fair to moderate) predictive performance of the machine learning models. Amongst the articles that compared the performance of conventional linear models with machine learning models, most found the machine learning models to perform better for predicting the outcomes [21,22,27], while one article found the conventional method to perform equally well, compared to the machine learning methods [23]. Despite the above chance predictive performance reported in most articles, the limitations posed by the small size of training datasets used to develop the models and the lack of external validation datasets has been widely acknowledged [18,21,25,34].

3. Discussion

Our review identified 15 articles focusing on the utilisation of PROMs for predicting outcomes by leveraging the analytical abilities of machine learning methods. Over the last decade, machine learning methods have received more attention in clinical research and are increasingly being adopted for furthering research in clinical analysis, modelling and building decision support systems for practitioners. The included articles presented promising research, demonstrating that as more and more healthcare data become available for developmental research, personalised treatment and medicine become more feasible with the help of machine learning-based decision support systems. Mobile applications allowing faster collection of PROMs data, as shown by Rahman et al. [30,31], is a promising way to collect more data frequently as well as to utilise the collected data for further research and development. Thus, the application of machine learning methods on PROMs data for predicting patient-specific outcomes appears to be a promising avenue and warrants further research.

3.1. Gaps and Challenges

The lack of external validation and non-availability of datasets used in the majority of the articles pose a major gap in the data availability for machine learning research. To drive the field forward, access to and open research questions in suitable datasets is a prerequisite. Datasets that are both comprehensive, complete, and readily available for research purposes, such as machine learning model development, are rare. Such datasets can facilitate the external validation by researchers in different disciplines and potentially inter-disciplinary collaboration. In other medical domains, opening pre-processed and experiment-ready datasets have shown that they draw attention to machine learning researchers and practitioners to explore different methods and benchmark the results [38–40]. As for the sizes of the datasets, eight of the fifteen articles included in this review used training datasets with more than 1000 patients (see Table 1), highlighting the sparsity of decent sized healthcare datasets for machine learning modelling. Furthermore, data collected with a different intent originally cannot automatically be used for machine learning due to uncertain or missing informed consent from participants. Most datasets collected from patients requires their consent for the utilisation of their data for various other purposes, which may not have been foreseen at the time of data collection. This may limit the ways in which patient data can be stored, used or distributed as well as the scope of the data.

Explainability and trustworthiness of the machine learning models are important challenges when it comes to developing clinical decision support systems. While a lot of attention has been given to developing accurate machine learning models, it is crucial to build systems that are trustworthy and interpretable. The users of such systems, for example medical researchers or clinicians, should be able to interpret the output of the machine learning models. Interpretations can be facilitated either through visualisations or explanations. This is an important aspect for clinicians, as they can focus on addressing the medical concerns rather than struggling with comprehension of the system's results.

Moreover, inconsistency was observed in reporting the development of the machine learning models in the articles. Only six articles reported the essential aspects of machine learning model development, such as feature selection and hyperparameter tuning, whereas in nine articles, this was either unclear or not stated at all, which can limit the reproducibility of results and further research.

Despite the progress in the development of machine learning models aimed at facilitating informed decision-making, there is still some more progress needed before these tools can be used in clinical practice. Specifically, external validation on large datasets of specific cohorts and thorough evaluation of the prediction tools are necessary before these tools can be integrated in clinical practice.

3.2. Limitations

The limitations of this review were that it was not possible to perform a meta-analysis of the results in the included articles due to various reasons, including, but not limited to, the heterogeneous study design, data non-availability, and study results, as summarised in Table 1 and discussed in Section 2.10. Out of the fifteen articles included in the analysis, only four articles reported their data source (national registry datasets), and one article stated that their dataset may be available upon reasonable request. However, none of the datasets were available during this literature review process for a meta-analysis. Further, we acknowledge that the articles retrieved in this literature review include only those articles that were retrieved during our search and met the inclusion criteria. As stated in the inclusion criteria, we included only those articles that focus solely on PROMs.

4. Conclusions

In summary, this literature review resulted in two main findings. First, there has been an increase during recent years in applying machine learning methods in exploring PROMs datasets for predicting patient-specific outcomes. Second, although the included articles have reported promising results and improvements [21,23,28], the lack of data availability and inconsistent reporting of machine learning model development as well as the use of different evaluation metrics prevents effective results reproduction and comparison. To conclude, utilising machine learning methods on PROMs datasets have the potential for assisting in clinical decision making; therefore, further research focusing on thorough validation is needed.

Author Contributions: Conceptualisation, D.V. and K.B.; review, D.V.; formal analysis, D.V.; writing—original draft preparation, D.V.; writing—review and editing, K.B., P.J.M., D.V. All authors have read and agreed to the published version of the manuscript.

Funding: This work is funded by the Back-UP EU project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 777090.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analysed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

NR	Not Reported
PROMs	Patient-Reported Outcome Measures
EHR	Electronic Health Records
CV	Cross Validation
LASSO	Least Absolute Shrinkage and Selection Operator
ANOVA	Analysis of Variance
RoC	Receiver Operating Characteristic Curve
MCID	Minimal Clinically Important Difference
NIMH	National Institute of Mental Health
NHS	National Health Service
HRS	Health and Retirement Study

References

1. Kingsley, C.; Patel, S. Patient-reported outcome measures and patient-reported experience measures. *Bja Educ.* **2017**, *17*, 137–144. [CrossRef]
2. Bingham III, C.O.; Noonan, V.K.; Auger, C.; Feldman, D.E.; Ahmed, S.; Bartlett, S.J. Montreal Accord on Patient-Reported Outcomes (PROs) use series—Paper 4: Patient-reported outcomes can inform clinical decision making in chronic care. *J. Clin. Epidemiol.* **2017**, *89*, 136–141. [CrossRef]
3. Barry, M.J.; Edgman-Levitan, S. Shared decision making—The pinnacle patient-centered care. *N. Engl. J. Med.* **2012**, *366*, 780–781. [CrossRef]
4. Coronado-Vázquez, V.; Canet-Fajas, C.; Delgado-Marroquín, M.T.; Magallón-Botaya, R.; Romero-Martín, M.; Gómez-Salgado, J. Interventions to facilitate shared decision-making using decision aids with patients in Primary Health Care: A systematic review. *Medicine* **2020**, *99*, e21389. [CrossRef]
5. Sepucha, K.R.; Atlas, S.J.; Chang, Y.; Freiberg, A.; Malchau, H.; Mangla, M.; Rubash, H.; Simmons, L.H.; Cha, T. Informed, Patient-Centered Decisions Associated with Better Health Outcomes in Orthopedics: Prospective Cohort Study. *Med. Decis. Mak.* **2018**, *38*, 1018–1026. [CrossRef]
6. Jayakumar, P.; Di, J.; Fu, J.; Craig, J.; Joughin, V.; Nadarajah, V.; Cope, J.; Bankes, M.; Earnshaw, P.; Shah, Z. A patient-focused technology-enabled program improves outcomes in primary total hip and knee replacement surgery. *JBJS Open Access* **2017**, *2*, e0023. [CrossRef] [PubMed]
7. Giga, A. How health leaders can benefit from predictive analytics. In *Healthcare Management Forum*; SAGE Publications: Los Angeles, CA, USA, 2017; Volume 30, pp. 274–277.
8. ShahabiKargar, Z.; Khanna, S.; Good, N.; Sattar, A.; Lind, J.; O'Dwyer, J. Predicting Procedure Duration to Improve Scheduling of Elective Surgery. In *PRICAI 2014: Trends in Artificial Intelligence*; Pham, D.N., Park, S.B., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 998–1009.
9. Kargar, Z.S.; Khanna, S.; Sattar, A. Using prediction to improve elective surgery scheduling. *Australas. Med. J.* **2013**, *6*, 287. [CrossRef]
10. Devi, S.P.; Rao, K.S.; Sangeetha, S.S. Prediction of surgery times and scheduling of operation theaters in optholmology department. *J. Med. Syst.* **2012**, *36*, 415–430. [CrossRef] [PubMed]
11. Wong, D.; Oliver, C.; Moonesinghe, S. Predicting postoperative morbidity in adult elective surgical patients using the Surgical Outcome Risk Tool (SORT). *BJA Br. J. Anaesth.* **2017**, *119*, 95–105. [CrossRef] [PubMed]
12. Moonesinghe, S.R.; Mythen, M.G.; Das, P.; Rowan, K.M.; Grocott, M.P. Risk stratification tools for predicting morbidity and mortality in adult patients undergoing major surgery: Qualitative systematic review. *Anesthesiology* **2013**, *119*, 959–981. [CrossRef]
13. Marufu, T.C.; White, S.; Griffiths, R.; Moonesinghe, S.; Moppett, I.K. Prediction of 30-day mortality after hip fracture surgery by the Nottingham Hip Fracture Score and the Surgical Outcome Risk Tool. *Anaesthesia* **2016**, *71*, 515–521. [CrossRef]
14. Singal, A.G.; Mukherjee, A.; Elmunzer, B.J.; Higgins, P.D.; Lok, A.S.; Zhu, J.; Marrero, J.A.; Waljee, A.K. Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma. *Am. J. Gastroenterol.* **2013**, *108*, 1723. [CrossRef] [PubMed]
15. Mansell, G.; Corp, N.; Wynne-Jones, G.; Hill, J.; Stynes, S.; van der Windt, D. Self-reported prognostic factors in adults reporting neck or low back pain: An umbrella review. *Eur. J. Pain* **2021**, *25*, 1627–1643. [CrossRef] [PubMed]
16. Krismer, M.; Van Tulder, M. Low back pain (non-specific). *Best Pract. Res. Clin. Rheumatol.* **2007**, *21*, 77–91. [CrossRef] [PubMed]
17. Waljee, A.K.; Higgins, P.D.; Singal, A.G. A primer on predictive models. *Clin. Transl. Gastroenterol.* **2014**, *5*, e44. [CrossRef] [PubMed]
18. Andrews, J.; Harrison, R.; Brown, L.; MacLean, L.; Hwang, F.; Smith, T.; Williams, E.A.; Timon, C.; Adlam, T.; Khadra, H.; et al. Using the NANA toolkit at home to predict older adults' future depression. *J. Affect. Disord.* **2017**, *213*, 187–190. [CrossRef]
19. Wang, X.; Gottumukkala, V. Patient Reported Outcomes: Is this the Missing Link in Patient-centered Perioperative Care? *Best Pract. Res. Clin. Anaesthesiol.* **2020**. [CrossRef]

20. Baumhauer, J. Patient-Reported Outcomes—Are They Living Up to Their Potential? *N. Engl. J. Med.* **2017**, *377*, 6–9. [CrossRef]
21. Shi, H.Y.; Tsai, J.T.; Chen, Y.M.; Culbertson, R.; Chang, H.T.; Hou, M.F. Predicting two-year quality of life after breast cancer surgery using artificial neural network and linear regression models. *Breast Cancer Res. Treat.* **2012**, *135*, 221–229. [CrossRef] [PubMed]
22. Huber, M.; Kurz, C.; Leidl, R. Predicting patient-reported outcomes following hip and knee replacement surgery using supervised machine learning. *BMC Med. Inform. Decis. Mak.* **2019**, *19*, 3. [CrossRef]
23. Pua, Y.H.; Kang, H.; Thumboo, J.; Clark, R.A.; Chew, E.S.X.; Poon, C.L.L.; Chong, H.C.; Yeo, S.J. Machine learning methods are comparable to logistic regression techniques in predicting severe walking limitation following total knee arthroplasty. *Knee Surg. Sport. Traumatol. Arthrosc.* **2019**, *28*, 3207–3216. [CrossRef]
24. 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? *Clin. Orthop. Relat. Res.* **2019**, *477*, 1267–1279. [CrossRef] [PubMed]
25. Polce, E.M.; Kunze, K.N.; Fu, M.; Garrigues, G.E.; Forsythe, B.; Nicholson, G.P.; Cole, B.J.; Verma, N.N. Development of Supervised Machine Learning Algorithms for Prediction of Satisfaction at Two Years Following Total Shoulder Arthroplasty. *J. Shoulder Elb. Surg.* **2020**, *30*, e290–e299. [CrossRef] [PubMed]
26. 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? *Clin. Orthop. Relat. Res.* **2019**, *477*, 452. [CrossRef] [PubMed]
27. Kessler, R.C.; van Loo, H.M.; Wardenaar, K.J.; Bossarte, R.M.; Brenner, L.A.; Cai, T.; Ebert, D.D.; Hwang, I.; Li, J.; de Jonge, P.; et al. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Mol. Psychiatry* **2016**, *21*, 1366–1371. [CrossRef]
28. Chekroud, A.M.; Zotti, R.J.; Shehzad, Z.; Gueorguieva, R.; Johnson, M.K.; Trivedi, M.H.; Cannon, T.D.; Krystal, J.H.; Corlett, P.R. Cross-trial prediction of treatment outcome in depression: A machine learning approach. *Lancet Psychiatry* **2016**, *3*, 243–250. [CrossRef]
29. Chekroud, A.M.; Gueorguieva, R.; Krumholz, H.M.; Trivedi, M.H.; Krystal, J.H.; McCarthy, G. Reevaluating the efficacy and predictability of antidepressant treatments: A symptom clustering approach. *JAMA Psychiatry* **2017**, *74*, 370–378. [CrossRef]
30. 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. *J. Med. Internet Res.* **2018**, *20*, e12001. [CrossRef]
31. 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 Med. Inform.* **2019**, *7*, e15601. [CrossRef]
32. Nijeweme-d'Hollosy, W.; van Velsen, L.; Poel, M.; Groothuis-Oudshoorn, C.; Soer, R.; Hermens, H. Evaluation of Three Machine Learning Models for Self-Referral Decision Support on Low Back Pain in Primary Care. *Int. J. Med. Inform.* **2018**, *110*, 31–41. [CrossRef]
33. Schiltz, N.; Dolansky, M.; Warner, D.; Stange, K.; Gravenstein, S.; Koroukian, S. Impact of Instrumental Activities of Daily Living Limitations on Hospital Readmission: An Observational Study Using Machine Learning. *J. Gen. Intern. Med.* **2020**, *35*, 2865–2872. [CrossRef] [PubMed]
34. Wang, Y.; Hays, R.D.; Marcus, M.; Maida, C.; Shen, J.; Xiong, D.; Coulter, I.; Lee, S.; Spolsky, V.; Crall, J.; et al. Developing Children's Oral Health Assessment Toolkits Using Machine Learning Algorithm. *JDR Clin. Transl. Res.* **2020**, *5*, 233–243. [CrossRef]
35. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
36. Bleeker, S.; Moll, H.; Steyerberg, E.; Donders, A.; Derksen-Lubsen, G.; Grobbee, D.; Moons, K. External validation is necessary in prediction research:: A clinical example. *J. Clin. Epidemiol.* **2003**, *56*, 826–832. [CrossRef]
37. Steyerberg, E.W.; Harrell, F.E., Jr. Prediction models need appropriate internal, internal-external, and external validation. *J. Clin. Epidemiol.* **2016**, *69*, 245. [CrossRef]
38. Xu, J.; Tong, L.; Yao, J.; Guo, Z.; Lui, K.Y.; Hu, X.; Cao, L.; Zhu, Y.; Huang, F.; Guan, X.; et al. Association of sex with clinical outcome in critically ill sepsis patients: A retrospective analysis of the large clinical database MIMIC-III. *Shock* **2019**, *52*, 146. [CrossRef] [PubMed]
39. Wang, S.; McDermott, M.B.; Chauhan, G.; Ghassemi, M.; Hughes, M.C.; Naumann, T. Mimic-extract: A data extraction, preprocessing, and representation pipeline for mimic-iii. In Proceedings of the ACM Conference on Health, Inference, and Learning, Toronto, ON, Canada, 2–4 April 2020; pp. 222–235.
40. Feng, M.; McSparron, J.I.; Kien, D.T.; Stone, D.J.; Roberts, D.H.; Schwartzstein, R.M.; Vieillard-Baron, A.; Celi, L.A. Transthoracic echocardiography and mortality in sepsis: Analysis of the MIMIC-III database. *Intensive Care Med.* **2018**, *44*, 884–892. [CrossRef] [PubMed]