**REVIEW**

**Open Access**

# Using artificial intelligence to predict patient outcomes from patient-reported outcome measures: a scoping review

Zuzanna Wójcik[1]* , Vania Dimitrova[2] , Lorraine Warrington[3] , Galina Velikova[3,4] and Kate Absolom[3,5]

## Abstract

**Purpose**  This scoping review aims to identify and summarise artificial intelligence (AI) methods applied to patient-reported outcome measures (PROMs) for prediction of patient outcomes, such as survival, quality of life, or treatment decisions.

**Introduction**  AI models have been successfully applied to predict outcomes for patients using mainly clinically focused data. However, systematic guidance for utilising AI and PROMs for patient outcome predictions is lacking. This leads to inconsistency of model development and evaluation, limited practical implications, and poor translation to clinical practice.

**Materials and methods**  This review was conducted across Web of Science, IEEE Xplore, ACM, Digital Library, Cochrane Central Register of Controlled Trials, Medline and Embase databases. Adapted search terms identified published research using AI models with patient-reported data for outcome predictions. Papers using PROMs data as input variables in AI models for prediction of patient outcomes were included.

**Results**  Three thousand and seventy-seven records were screened, 94 of which were included in the analysis. AI models applied to PROMs data for outcome predictions are most commonly used in orthopaedics and oncology. Poor reporting of model hyperparameters and inconsistent techniques of handling class imbalance and missingness in data were found. The absence of external model validation, participants' ethnicity information and stakeholders involvement was common.

**Conclusion**  The results highlight inconsistencies in conducting and reporting of AI research involving PROMs in patients' outcomes predictions, which reduces the reproducibility of the studies. Recommendations for external validation and stakeholders' involvement are given to increase the opportunities for applying AI models in clinical practice.

*Correspondence:
Zuzanna Wójcik
sczw@leeds.ac.uk
[1] UKRI Centre for Doctoral Training in Artificial Intelligence for Medical Diagnosis and Care, University of Leeds, Leeds, UK
[2] School of Computer Science, University of Leeds, Leeds, UK
[3] Leeds Institute of Medical Research, University of Leeds, St James's University Hospital, Leeds, UK
[4] Leeds Cancer Centre, Leeds Teaching Hospitals NHS Trust, Leeds, UK
[5] Leeds Institute of Health Sciences, University of Leeds, Leeds, UK

## Introduction

Artificial Intelligence (AI) is a field of computer science and engineering which uses computer systems able to mimic intelligent behaviour [1]. AI is known to have potential to improve the effectiveness, accessibility and accuracy of screening, diagnosis and treatment in many areas of health [2, 3]. AI models predicting patient outcomes can achieve high performance, and as a result aid clinical decisions and improve quality of healthcare [3]. AI has been applied to various data types in medicine,

Wójcik *et al. Health and Quality of Life Outcomes*        (2025) 23:37

Page 2 of 15

using mainly clinical data, such as diagnostic images, genetic data, or brain activity data [4].

While there is a growing attention at patient-reported data in clinical practice and some attempts to use AI models on such data exist [5], systematic guidance on how to apply AI on patient-reported data for outcome predictions is lacking. Patient-reported data can be collected using patient-reported outcome measures (PROMs). These are questionnaires which measure patients' perception on their health status, without being influenced by clinical opinion [6]. PROMs data can be either standardised and validated tools designed to capture patients' reports, or any other forms of symptom and quality of life measures [7]. For instance, mobile applications for PROMs collection, have been widely used in healthcare and have potential to improve the quality and personalisation of patient care [8]. The recent systematic evaluation of PROMs in clinical trials of AI health technologies has shown that patients' perspective is central even in novel technological advancements [9].

Unfortunately, the complexity of PROMs data and limited universal guidelines for AI use in healthcare research [10], can lead to inconsistent reporting of study design and evaluation [11]. Furthermore, studies often lack reproducibility, external validity [12], and generalisability of the results to the clinical context [10]. Inadequate and inconsistent selection of patient-reported input data also introduces a challenge to useful application of patient-centred AI models in healthcare [13]. Additionally, there is a lack of patient and clinician involvement in the process of study design, which plays an important role in addressing bias in AI research for healthcare [14].

There are existing literature reviews exploring AI models applied on PROMs data. For example, a scoping review from 2021 investigated PROMs as standalone input variables in models, however, they did not explore reproducibility and clinical adoption of studies. Moreover, only 2 medically oriented literature databases were searched, while databases from engineering and computer science backgrounds were not considered [5]. Other existing reviews focused on specific healthcare domains (e.g. oncology) and did not investigate all potential applications of using AI and PROMs data. [15–17].

This review aims to address the gap in the literature by investigating AI models used in primary studies for predicting patient outcomes using PROMs. It focuses on methodological rigour of conducting, evaluating and reporting AI research including PROMs as input data. It highlights the importance of ensuring standardised dataset description and justification for chosen methods of model development and evaluation, focusing on clinical relevance. Recommendations for engaging stakeholders, including patients, are suggested.

## Materials and methods

The methodology of this scoping review was based on the Joanna Briggs Institute (JBI) guidance [18]. The review protocol is available on Open Science Framework [19]. The completed PRISMA checklist for scoping reviews [20] is added in the Supplementary Materials Fig. 1 and 2.

### Search strategy

The databases used to search for relevant papers were: Web of Science, IEEE Xplore, ACM Digital Library, Cochrane Central Register of Controlled Trials, Medline and Embase. These databases were selected to include the variety of fields publishing papers on AI in medicine, covering both medical and engineering aspects. The keywords adapted to each database are listed in Supplementary Materials Table 1. Initially, the limited search of Web of Science and Medline was conducted to analyse and approve the keywords. The finalised search of all the records across all the databases was completed on the 7th November 2023. The reference lists of all relevant reports were also screened. All studies identified through the search strategy were exported to Endnote citation management system.

### Inclusion and exclusion criteria

The inclusion and exclusion criteria followed the Population/Concept/Context (PCC) framework [18] and are described in Table 1. The participants in the papers included in the review were patients, whose symptoms and quality of life data were recorded using various PROMs. These can include mobile applications, or surveys completed either online or in a clinic. The type of data can be collected through either validated and widely used PROMs, or any other patient self-reports. Papers reporting the use of PROMs data as both a predictive and predicted variable were included in the study. If PROMs data were only used as a predicted variable, and not included as inputs, the reference was excluded. The concept was the methods of AI used for predictions of the patients' outcomes. Papers that explicitly mentioned use of AI or Machine Learning methods were included. Any papers using AI models for purposes other than prediction were excluded. Papers that reported prediction of patient outcomes in the healthcare context were included in the analysis. These outcomes should belong to the categories of patient-related outcomes identified by Kersting et al. (2020) [21], presented in Table 1. The broad understanding of healthcare context allowed focusing on the AI used for various medical reasons.

### Review process

All duplicates found in the databases were removed automatically in Endnote. Titles and abstracts of the

**Table 1** Inclusion and Exclusion criteria for the study selection

| Category | Inclusion criteria | Exclusion criteria |
|---|---|---|
| Input variables | Data reported by patients using standardised PROMs; electronic data collection designed for self-reporting of symptoms or clinical outcomes. These can be used with combination of different non-patient reported data | Only not patient-reported data, e.g., data reported through a clinician, recorded, written down during an appointment, data reported on online forums/social media, or data from physiological measurements |
| Output variables | Any data describing patient outcomes: sleep behaviour, coping and self-efficacy, healthcare utilisation, body image perception, function, communication skills, reliability of diagnosis and therapy, optimal support, confidence in therapy, satisfaction, cognitive performance, treatment decision, disease control, daily activities, reoperation, mental health, quality of life, mobility, co-morbidities, pain, survival, adverse events, and symptoms | Output that does not relate to health or healthcare, or papers which did not aim to predict outcomes |
| Models | Machine learning or deep learning predictive models | Statistical models (e.g., only regression analysis) |
| Paper type | Primary research reported in English language | Abstracts only, theses, dissertations, letters to editors, guidelines, commentaries, introductions, papers published not in English, or review papers |

papers were screened by a researcher and re-selected based on inclusion and exclusion criteria, presented in Table 1. Full texts of articles admitted to the study were assessed against the inclusion and exclusion criteria again. The researcher's approach was validated through a second reviewer, who repeated scanning through 10% of abstracts, selected full texts and compared their results with the first reviewer. The validation showed high consistency between the reviewers' decisions, as out of 218 validated papers, 184 (84.4%) were consistently selected or rejected. Therefore, no further validation was performed.

### Data extraction and analysis

The data was extracted from all papers selected for this review. Extracted and summarised information for each included paper is presented in Supplementary Materials Tables 1 and 2. A second reviewer extracted data from 10% of admitted papers for the purpose of validation, and the extracted information was compared and agreed between the reviewers. The summary of data was reported in tabular form in Excel spreadsheet and presented in a narrative form in this review.

The extracted and analysed data included:

- Study characteristics (country and year of publication, healthcare domain, input PROMs variables used, types of PROMs, output variable types, and sample sizes)
- Data pre-processing (missingness in the datasets, missing data imputation techniques, class distribution, techniques for handling class imbalance)

- Model development (types of AI models used, frequency of AI models used, AI techniques for addressing temporality in data, hyperparameter tuning)
- Model evaluation (performance metrics used, variable importance, best-performing AI models)
- Clinical relevance and adoption (patients and clinicians involvement in the study design, validation and deployment stage of research, reporting of sociodemographic information)

### Results

Out of 3077 records screened, 94 were selected for analysis in this review. PRISMA diagram [20] (Fig. 1) illustrates the process of paper selection. The reasons for paper exclusions were: no full-text available (38.1%), no PROMs used as input variables (35%), or no AI models used (12%), or methods did not aim to predict patient outcomes (8.75%).

### Study characteristics

Among the identified studies, 33 (35%) were conducted in USA, 31 (33%) in Europe, 6 (6%) in Canada, 15 (16%) in Asia, 1 (1%) in South America, 1 (1%) in New Zealand, and 1 (1%) in Turkey. Six (6%) studies were conducted internationally (USA and Canada ($n= 2$, 2%); UK and USA ($n= 2$, 2%); Canada and Sweden ($n= 1$, 1%); Europe, US, Australia and Israel (n =1, 1%)). Identified papers focused on orthopedics ($n= 38$, 40%), oncology ($n= 22$, 23%), mental health ($n= 17$, 18%), respiratory ($n= 8$, 9%), neurology ($n= 4$, 4%) and other domains ($n= 5$, 5%), which appeared only once: hearing, endometriosis, palliative care, sub-health state, and cardiovascular. The studies were published between 2010 and 2023 (Fig. 2). The
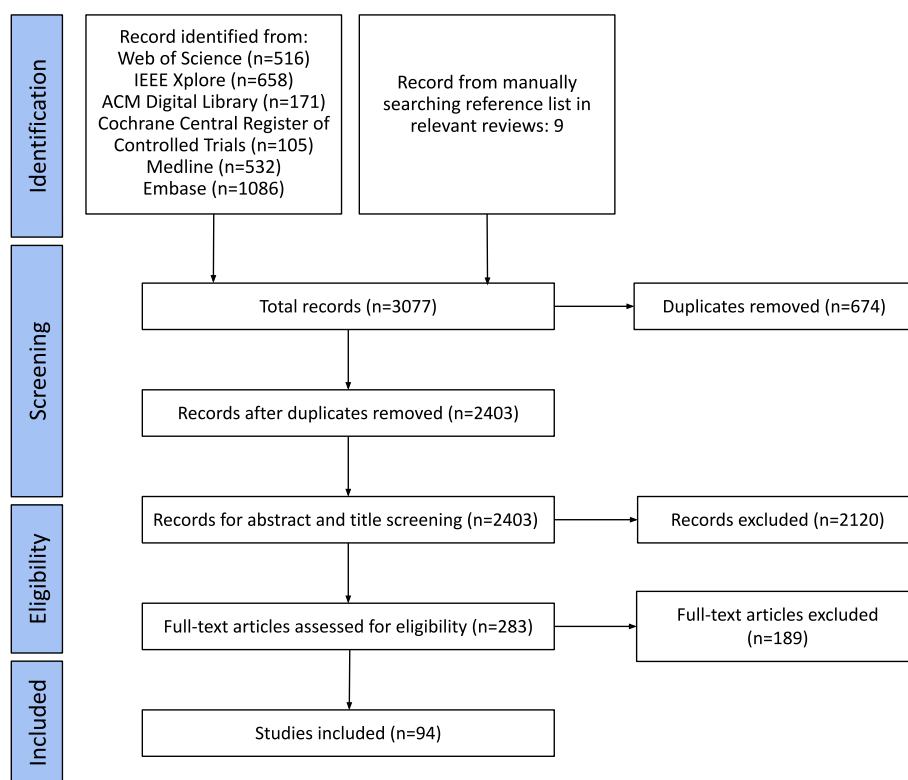
**Fig. 1** PRISMA flow diagram

data were obtained either from existing registry/database ($n=44$, 47%), or pre-existing or current research studies ($n=47$, 50%, not reported: $n=3$, 3%).The self-reported input variables were combined with clinical and demographic data ($n=63$, 67%), only demographic data ($n=14$, 15%), only clinical data ($n=3$, 3%), or other types of data ($n=4$, 4%), such as wearable, electroencephalographic, bio-mechanical, or family data. Ten studies used only self-reported data for predictions. Most papers ($n=63$, 67%) were predicting self-reported outcomes, 14 (15%) of which were Minimally Clinically Important Differences (MCID) between pre- and post-clinical event data collection. Other papers used either only objectively measured outcome ($n=28$, 30%), or a combination of self-reported and objective outcomes ($n=3$, 3%). Sample sizes of the papers varied from 20 to 1,434,868 (mean $=25,888$, median $=1022$, 1st quartile $=429.75$, 3rd quartile $=2879.75$). The quartiles do not indicate clear boundaries between the data, as there are small differences in the sample sizes. Hence, a boundary-based approach was followed instead of quartile-based: very small ($<300$), small (300–700), medium (701–2000), large (2001–20000), and very large ($>20,000$), as presented in Fig. 3. The vast majority of studies ($n=69$, 73%) used condition-specific PROMs, such as orthopedic-specific Knee Injury

and Osteoarthritis Outcome Score (KOOS) [22] or cancer-specific EORTC Core Quality of Life Questionnaire (QLQ-C30) [23]. In 31 papers (33%) condition-specific measures with generic questionnaires, for example Euro-Qol- 5D (EQ- 5D) [24]. Twelve papers (13%) used generic measures only. Out of 81 (86%) papers that reported the types of questionnaires used, 18 (22%) focused on physical health, 11 (14%) on mental health, and 52 (64%) on both.

**Data pre-processing**
Thirty papers did not report the missingness in the dataset. Therefore, it is uncertain if the datasets in these studies did not have any missing data, or if missing data were not disclosed. Out of 64 papers (68%) which reported missing data, only 1 (2%) stated that there was no missingness in the dataset. Ten papers (16%) which reported having missing data did not report how the missingness was handled or addressed. Out of all papers, only 53 (56%) reported the technique for data imputation (Fig. 4c). The 2 most common techniques were complete case analysis ($n=16$, 30%), and mean/median/mode imputation ($n=15$, 28%). Most papers ($N=89$, 95%) used classification as a prediction method. Fourteen (16%) of these did not provide any information about the class
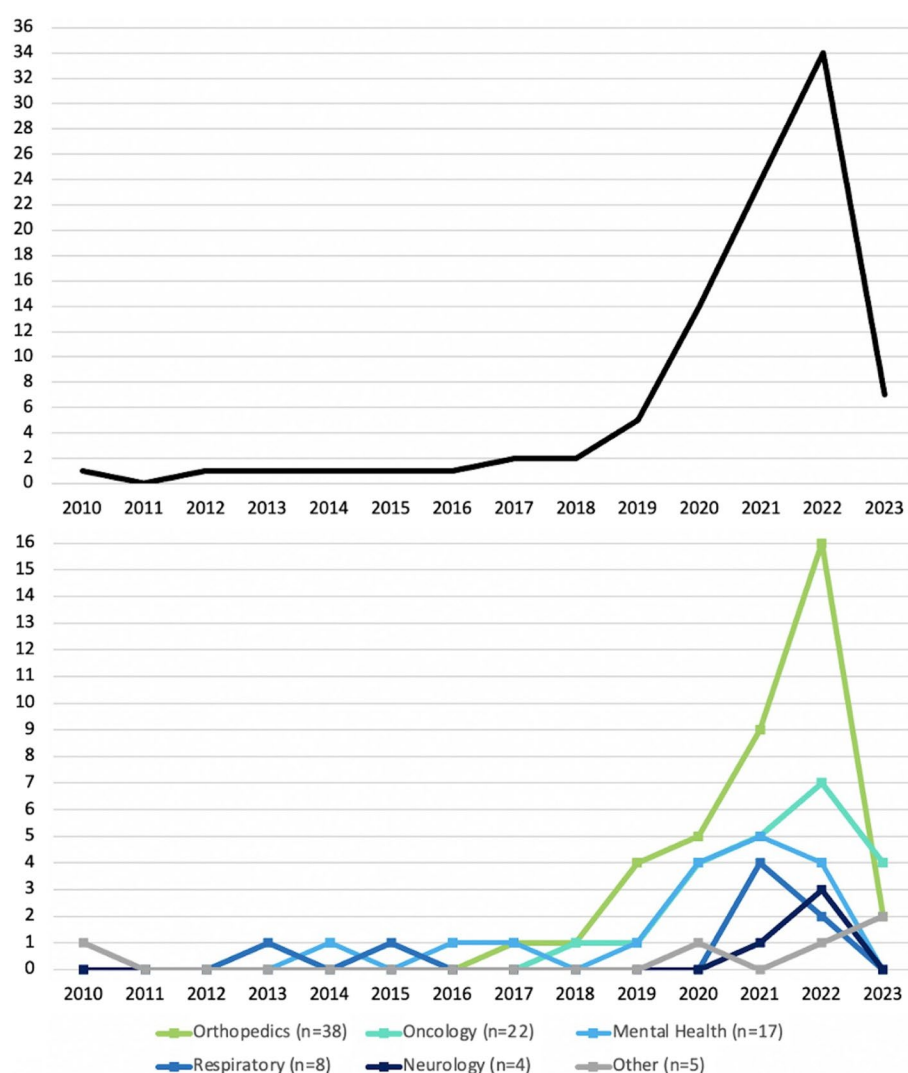
**Fig. 2** Year of publication of all 94 studies (top figure) and studies based on health domain (bottom figure)

distribution. All papers which reported class distribution ($n = 75$, 80%) performed binary classification. Out of these papers, only 11 (15%) had balanced classes (maximum imbalance ratio of 60:40 between the minority and majority class [25]). Sixty-four papers (68%) used dataset with imbalanced classes, 29 (45%) of which did not mention the class imbalance problem. Thirty-five papers (55%) acknowledged the issue but 13 (37%) of them left the data imbalanced. In total, 22 papers (23%) reported the need for balancing the classes, but there was inconsistency in the methods across papers (Fig. 4).

**Model development**

Most papers ($n = 84$, 89%) used multiple AI models for outcomes prediction. Forty papers (43%) used only traditional machine learning (ML) models, 5 (5%) only deep learning (DL) models, and 49 (52%) both ML and DL. The most frequently used models were regression models ($n = 61$, 65%), including linear, logistic, ridge and LASSO regressions; boosting methods ($n = 53$, 56%), including adaptive boosting, extreme gradient boosting and gradient boosting machine; random forest ($n = 50$, 53%); artificial neural network ($n = 43$, 46%), including single-or multi-layer perceptrons; and support vector machine ($n = 39$, 41%) (Fig. 5). Most studies ($n = 74$, 81%) applied AI models on data recorded in one time-point. The remaining studies trained their models on data collected in multiple time-points (Table 2). Out of these, 3 studies (3%) reported using models that process the temporal dependencies in the data, such as long-short term memory (LSTM) model [26, 27], and recurrent neural network with gated recurrent units (GRU) [28]. Nine studies (12%) considered temporality through coding it in the feature sets, and 5 papers (7%) did not address
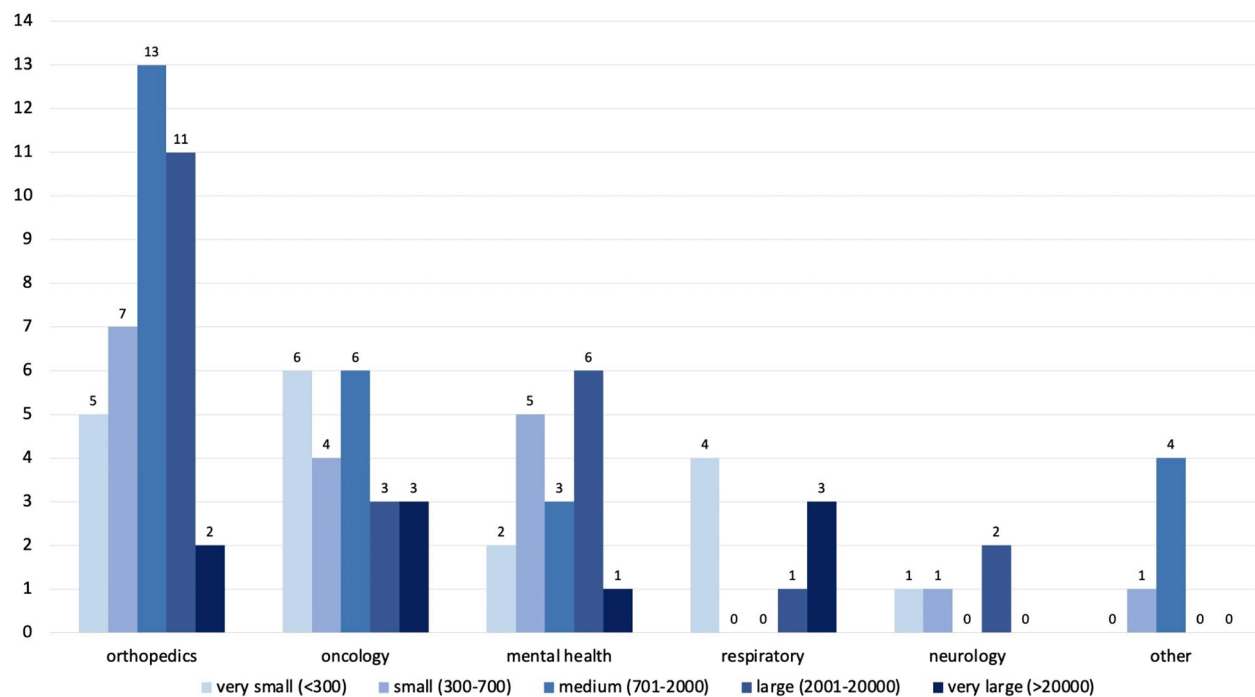
**Fig. 3** Number of papers categorised based on sample sizes in each healthcare domain
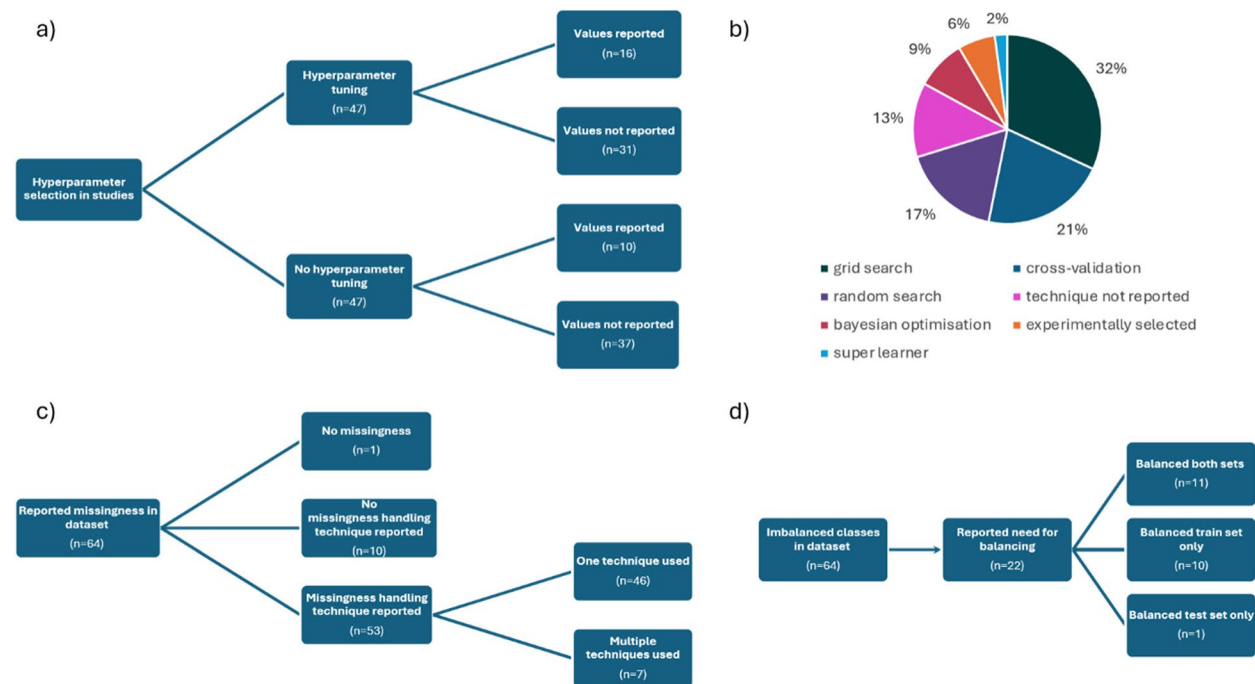


**Fig. 4** Reporting of pre-processing and model development methods in the studies. Sub-figure **a**) Frequency of hyperparameter tuning and values reporting. Sub-figure **b**) Proportion of hyperparameter tuning techniques. Sub-figure **c**) Missingness reporting and imputation in papers. Sub-figure **d**) Handling class imbalance in studies
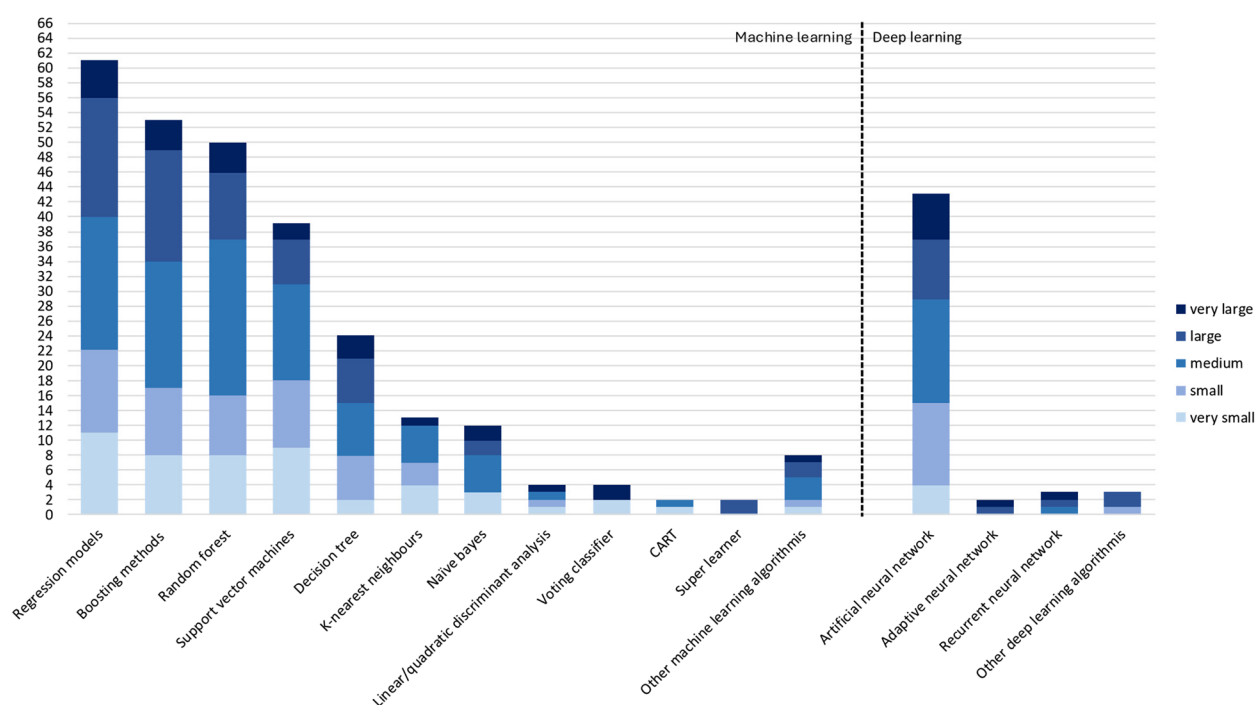
**Fig. 5** Frequency of algorithms used on datasets with very small (fewer than 300), small (300–700), medium (701–2000), large (2001–20000) and very large (more than 20,000) sample size

temporality at all (Table 3). Half of the papers in this review (*n* = 47, 50%) reported performing hyperparameter tuning, and out of these only 16 (34%) reported used hyperparameters (Fig. 4a)).

**Model evaluation**

The evaluation metrics varied across the studies. Area under the curve (AUC) was most commonly used (*n* = 60, 64%), and 32 (53%) of papers used this value to assess model performance with imbalanced classes. Other frequently used performance metrics were recall, also known as sensitivity (*n* = 44, 47%), accuracy (*n* = 43, 46%), and specificity (*n* = 34, 36%). The majority of the studies used multiple performance metrics (*n* = 83, 88%). Variable importance analysis was performed by 64 studies (68%) 61 of which (95%) reported PROMs data being valuable for prediction. Seventy-nine papers (84%) provided information on the best performing model. Regression models were the most frequently selected as best-performing algorithms (*n* = 24, 30%), followed by boosting methods (*n* = 20, 25%), random forest (*n* = 10, 13%) and neural network (*n* = 10, 13%).

**Clinical relevance and adoption**

No studies reported that the developed methods had been applied in the clinical practice. Although we acknowledge, that in such multidisciplinary research

clinicians are generally involved in the study design, only 3 papers (3%) explicitly mentioned how clinicians contributed to the model development. They helped selecting input variables [30, 111], or creating a testing set [100]. No papers mentioned patients involvement in the model development or any part of the study design. The majority of papers reported age (*n* = 63, 67%) and gender (*n* = 60, 64%) of study participants and only 24 (26%) studies reported ethnicity. Papers were classified into 3 different categories, inspired by a previously conducted scoping review [118]: internal validation (one source of data used for training and validation, including cross-validation or holdout sample for validation on unseen data from the same dataset), external validation (the model developed on one dataset and then tested/validated on a completely new (i.e. external) dataset) or deployment ("integrated into a prototype application, and evaluated for its feasibility in clinical workflows"[118]). Based on these definitions, 81 papers (86%) were in the internal validation stage, 10 (11%) completed external validation, and 3 papers (3%) were in the deployment stage.

**Discussion**

This scoping review aimed to identify AI methods used on PROMs data to predict patient outcomes. The analysis of 94 papers allowed the exploration of algorithms applied on complex patient-reported data and revealed

Wójcik *et al. Health and Quality of Life Outcomes*     (2025) 23:37

Page 8 of 15

**Table 2** Machine learning and deep learning models used on data collected in one and multiple timepoints, ordered by number of publications

| Machine Learning (ML) | One Timepoint | Multiple Timepoints |
|---|---|---|
| Regression (*n* = 61) | [29–39]<br>[40–50]<br>[40, 51–60]<br>[61–71]<br>[72–75]<br>[76] | [77–86]<br>[28, 87, 88] |
| Boosting (*n* = 53) | [31, 32, 36–40, 42, 44, 89, 90]<br>[45, 48, 49, 51–53, 91–94]<br>[40, 56, 58, 60, 62–64, 95–97]<br>[2, 33, 66, 69, 72, 73, 75, 98–100]<br>[70, 74, 101]<br>[76]<br>[102] | [28, 77–79, 81–84, 103, 104] |
| Random Forest (*n* = 50) | [29–33, 36, 41, 42, 44, 89, 105]<br>[45–48, 50–53, 56, 92, 106]<br>[58, 60, 61, 64, 65, 69, 75, 95, 100, 107, 108]<br>[62, 63, 71, 73, 74, 97, 101] | [26, 28, 77, 78, 80, 87, 88, 104, 109, 110] |
| Support Vector Machine (*n* = 39) | [29, 31, 32, 34, 37, 41–43, 45, 46, 48]<br>[26, 50–52, 56, 57, 92, 96, 111, 112]<br>[61, 62, 64, 66, 71, 74, 89, 107, 108]<br>[102] | [78, 81, 83, 84, 87, 88, 103, 104] |
| Decision Tree (*n* = 24) | [37, 39, 41, 57, 92–95, 111, 112]<br>[2, 58, 63, 74, 96, 98, 100, 101, 108] | [77, 79, 84, 103, 109] |
| K-Nearest-Neighbours (*n* = 13) | [34, 36, 37, 39, 48, 56, 61, 74, 113] | [26, 83, 87, 104] |
| Naïve Bayes (*n* = 12) | [36, 44, 50, 57, 70, 73, 74, 94, 108] | [26, 104, 109] |
| Voting Classifier (*n* = 4) | [37, 44] | [28, 82] |
| Discriminant Analysis (*n* = 4) | [34, 38, 94, 108] | None reported |
| Classification and Regression Tree (*n* = 2) | [34, 46] | None reported |
| Super Learner (*n* = 2) | [62, 69] | None reported |
| Other ML Methods (*n* = 8) | Wide and Deep [49]<br>Stochastic Gradient Descent [61] Bagging [63]<br>Bayesian Updating Algorithm [66]<br>Graphical Gaussian Model [67]<br>Multivariate Adaptive Regression Spline [41] | Hierarchical Gaussian Process [85]<br>Autoregressive Integrated Moving Average [26] |
| **Deep Learning (DL)** | **One Timepoint** | **Multiple Timepoints** |
| Multilayer Perceptron (*n* = 43) | [29, 30, 35–37, 39–42, 45, 90]<br>[48, 52–54, 56, 92, 94, 111, 112, 114]<br>[40, 57, 59, 61, 64, 66, 68, 96, 107, 115]<br>[72–75]<br>[76] | [26, 28, 79, 83, 84, 86, 104, 109, 116] |
| Recurrent Neural Network (RNN) (*n* = 3) | None reported | Long-Short Term Memory [26, 27]<br>RNN with Gated Recurrent Units [28] |
| Other DL Methods (*n* = 4) | Adaptive Neural Network [36]<br>Stacking Algorithm [57]<br>Bayesian Network Model [117] | Adaptive Neural Network [27] |

In the square brackets we list the number of the cited paper, according to the reference lis

the opportunities, challenges and best practice recommendations for AI medical research involving PROMs. The main findings suggest the variety of data types and evaluation metrics used, as well as inconsistencies in data pre-processing and model development design and reporting.

## Study characteristics

Due to fragmented data collection of PROMs, incorporating them into AI systems is very challenging [119]. Therefore, the majority of papers in this study have small sample size. In orthopedics settings PROMs have been increasingly collected as a part of routine care [120], which explains the large proportion of orthopedics papers with medium-to-large datasets. The large sample size was common in mental health papers, as

Wójcik *et al. Health and Quality of Life Outcomes*      (2025) 23:37

Page 9 of 15

**Table 3** Methods of addressing temporality in time-series data

| Level of addressing temporality | Method of addressing temporality | Description of the method | Sample size | Frequency |
|---|---|---|---|---|
| Addressed by the model | Recurrent Neural Network (RNN) | Data transformed to 3D array and fed in the LSTM model [26] Events encoded by Adaptive Net, pooled by LSTM model [27] | 823<br>9,500 | Weekly<br>Irregular[a] |
| | | RNN with GRU, considering each treatment as timestep [28] | 105,129 | Irregular[a] |
| Addressed in the features | Measured change | Change of measurement from baseline [84]<br>Change in mean measurements from baseline [88]<br>Mean daily change from the 24-h baseline period [87]<br>Change in symptom severity from previous report [78] | 245<br>31,700<br>116<br>34 | Every 90 days<br>Daily<br>Twice a day<br>Weekly |
| | Binary outcome | Variable indicated if a report is followed by exacerbation event [80] Occurrence of symptom in any day of a time window [85] | 2,374<br>182,991 | Daily<br>Daily (3 days) |
| | | Dichotomised score one week following the prediction date [104] | 210 | Weekly |
| | Feature for each timeline | Score added as an input feature at every measurement [82] Created a timeline of best overall responses (BORs) [103] | 83<br>31 | Weekly<br>Weekly |
| Not considered | Model for each timeline | Treated the 2- and 8-week measures as if assessed at baseline [81] Three models that used 7, 14, and 21 days as inputs [116] | 1,003<br>20 | 3 time-points<br>3 time-points |
| | Selected 1 value for analysis | If patient had multiple follow-up events, the first was chosen [77] The assessment with the highest overall score was used [79] | 494<br>11,761 | Irregular[a] Bi-weekly |
| | | Score was updated at each assessment [86] | 212,615 | Irregular[a] |

[a] Irregular measurements indicate that the reports were completed at any clinical event that occurred

Three papers which used time-series data did not report how the temporality was addressed [83, 109, 110], and are not included in this table

mental health screening and diagnostic tools are usually based on PROMs, and there is a long-standing history of using such tools [121]. Contrary to orthopedics and mental health settings, PROMs collection in other healthcare domains is very limited. The respiratory datasets were mainly  very small or very large. The papers with very large sample size were predicting outcomes related to COVID- 19 pandemic, where mobile applications collecting PROMs became more common [122]. Most of the studies analysed data collected specifically for research studies, rather than in clinical practice, which might introduce biases related to inclusion and exclusion criteria. Due to inconsistent PROMs questionnaires used across different studies, the comparison of results is limited. Therefore, using standardised and validated measures can help explore the overall predictive value of PROMs. The peak in the use of AI methods for all domains was between 2021 and 2022. This recent increase is compatible with a scoping review on AI in healthcare, where 71% of studies were published between 2020 and 2022 [123].

**Data pre-processing**

Missing data in AI research is an important aspect to investigate, as it can lead to various biases [124]. Therefore, the inconsistencies in reporting data quality in the analysed studies are concerning. The justification of using data imputation techniques was also poor, whilst most commonly used techniques (complete case analysis and mean/median/mode imputations) can frequently cause bias [124]. Only a small number of papers used KNN-based imputation, which can reach the accuracy of complete data with a low performance difference [125]. Studies applying AI methods on PROMs  should ensure that missing data are reported and any imputation methods are justified [124]. Another inconsistency between the papers was caused by various methods for handling class imbalance in classification tasks. Papers

which reported class imbalance often did not attempt to balance the data, which prevents models from appropriate learning from the training set. Furthermore, most of these papers used AUC as a performance metric, which require the balanced setting to avoid bias [126]. The papers which reported balancing data have also done it inconsistently and without justification. Balancing data prior to train and test split can cause issues in model validation, as real data is never perfectly balanced. Therefore, it is important to evaluate model performance on test set unaffected by sampling methods [127]. The choice of performance metrics should also be justified and able to uncover potential bias caused by class imbalance (for example balanced accuracy and F1 score, instead of accuracy and AUC).

### Model development

The studies in this review reported model development process inconsistently, with majority of studies missing model hyperparameters reporting. According to Jha et al. (2023)[128], it should be"the ethical obligation" to document all stages of model development that are essential for the reproducibility of results. Therefore, model hyperparameters and their optimisation technique should always be reported and justified. The missingness in data was also handled and reported inconsistently, which is an important step for reproducibility as well. The lack of large PROMs datasets prevents applying deep learning methods, which can be extremely useful in capturing patterns in high-dimensional data or dependencies that other algorithms can't capture [129]. Only the simple"vanilla" neural network was applied more often than some of the basic ML models. Studies which collected data in multiple timepoints often did not address the temporality at all or analysed time-series data through data pre-processing strategies and conventional ML models. Only 3 papers used DL models that are appropriate for temporal processing. These are for example LSTM or GRU methods. Most papers in this review chose various ways to include temporal information through feature engineering, as described in Table 3. Nevertheless, DL models have been more successful in accurate predictions of patients' outcomes when applied to time-series data, as they are able to process more complex dependencies in high dimensionality and temporality of medical data [130].

### Model evaluation

Most studies used multiple evaluation metrics, which allow between-studies comparisons and in-depth analysis of model performance. Variable importance analysis was also commonly conducted, which supports the explainability of AI models [131]. Furthermore, the studies used multiple models, which allowed them to select the best-performing one. The analysis of these showed that most common models were rarely selected as best models (e.g., random forest was selected as the best model only 20% of the times). However, voting classifier was used only 4 times, but selected as the best model 3 times. This suggest that further studies should perhaps pay more attention to models that are used less-frequently, which have a potential to perform better.

### Clinical relevance and adoption

Limited reporting of clinician and patient engagement in the study development is of concern. This process is known as crucial for ensuring a patient-centred research and feasibility of the studies [132]. Involving stakeholders also can help building trust of the public to AI researchers, and as a result support the implementation of the studied tools in clinical practice [133]. Another issue arising from this review is the lack of external validation of the model performance in the studies, which is an essential step to potential clinical adoption. Assessing the model performance in a different setting may show different model performance which might suggest bias in the original study [118]. This review shows that new models keep being developed to address original problems, without taking the studies further and exploring their potential in the real-world settings. Therefore, validating existing models on external datasets and communicating the design and results with stakeholders should be the next step to support the adoption of AI methods in clinical practice. Furthermore, the majority of papers did not provide any information on the ethnicity of the study participants. Ensuring diverse study population is an essential ethical consideration and lack of ethnic information can further contribute to deepening healthcare inequalities [128].

### Strengths and limitations

The main strength of this review is that it identifies AI models applied on self-reported data for predicting patient outcomes in all healthcare domains. The use of 6 different databases from both health and computer science field helped reaching many relevant papers, which might have been omitted by reviews using limited number of databases. This paper also analyses the rigour of model development and evaluation reporting. It focuses on clinical adoption potential, from the perspective of patients and clinicians involvement and ethical consideration of participants' diversity. The limitations of this study include the possibility of omitting the studies published in language other than English. Only published studies were considered, which might affect the conclusions and further deepen the publication bias. Since the

Wójcik *et al. Health and Quality of Life Outcomes*     (2025) 23:37

Page 11 of 15

study focuses on rigour in model development, evaluation, and wider stakeholder engagement, it is important to note, that the results are based on what was reported, and not what was done in the included papers.

## Conclusions

The analysis of 94 papers in this scoping review revealed the potential of using PROMs data in AI healthcare research, and inconsistencies in conducting and reporting these studies. It showed the importance of justification of chosen data pre-processing and model development methods, and the involvement of all stakeholders during the study. Our future work will involve applying AI on PROMs data and further explore the potential of time-series patient-reported data for healthcare outcomes predictions. We believe that insights from this paper can inform the rigorous implementation of AI models in clinical practice.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12955-025-02365-z.

Supplementary Material 1. Figure 1: PRISMA checklist for scoping reviews part 1. Figure 2: PRISMA checklist for scoping reviews part 2. Table 1: Search strategy. Table 2: Study characteristics and pre-processing methods used by studies included in the review. Table 3: Model development and evaluation, including study characteristics of papers included in the review.

## Authors' contributions
Zuzanna Wójcik: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing—Original Draft, Writing—Review & Editing, Visualization, Project administration Vania Dimitrova: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Writing—Original Draft, Writing—Review & Editing, Visualization, Supervision, Funding Acquisition Lorraine Warrington: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing—Original Draft, Writing—Review & Editing, Supervision Galina Velikova: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Writing—Original Draft, Writing—Review & Editing, Supervision, Funding Acquisition Kate Absolom: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Writing—Original Draft, Writing—Review & Editing, Supervision.

## Declarations

### Ethics approval and consent to participate
Not applicable.

### Competing interests
Galina Velikova: Honoraria: Pfizer, Novartis, Eisai, Lilly Advisory boards: Consultancy fees from AstraZeneca, Roche, Novartis, Pfizer, Seagen, Eisai, Sanofi April 2024 = AZ working group (unpaid) Institutional grant: Pfizer.

## References
1. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2019;9(4): e1312.
2. P. Rajpurkar, J. Yang, N. Dass, V. Vale, A.S. Keller, J. Irvin, and L.M. Williams Evaluation of a machine learning model based on pretreatment symptoms and electroencephalographic features to predict outcomes of antidepressant treatment in adults with depression: a prespecified secondary analysis of a randomized clinical trial. JAMA network open, 3(6),e206653-e206653. 2020
3. Yang CC. Explainable Artificial Intelligence for Predictive Modeling in Healthcare. Journal of Healthcare Informatics Research. 2022;6:228–39.
4. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, Wang Y, Dong Q, Shen H, Wang Y. Artificial intelligence in healthcare: past, present and future. Stroke and Vascular Neurology. 2017;2:230–43.
5. D. Verma, K. Bach, and P. J. Mork, "Application of Machine Learning Methods on Patient Reported Outcome Measurements for Predicting Outcomes: A Literature Review," Informatics, vol. 8, p. 56, 2021. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.
6. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. BJA Education. 2017;17:137–44.
7. D. Denny, S. B. Nutakki, K. Alston, and M. Markman, "A review of patient self-reported symptom data with a focus on pain," Journal of Clinical Oncology, vol. 32, pp. 79–79, 2014. Publisher: Wolters Kluwer.
8. Awad A, Trenfield SJ, Pollard TD, Ong JJ, Elbadawi M, McCoubrey LE, Goyanes A, Gaisford S, Basit AW. Connected healthcare: Improving patient care using digital health technologies. Adv Drug Deliv Rev. 2021;178: 113958.
9. F. J. Pearce, S. C. Rivera, X. Liu, E. Manna, A. K. Denniston, and M. J. Calvert, "The role of patientreported outcome measures in trials of artificial intelligence health technologies: a systematic evaluation of ClinicalTrials.gov records (1997–2022)," The Lancet Digital Health, vol. 5, pp. e160–e167, 2023. Publisher: Elsevier.
10. B. khan, H. Fatima, A. Qureshi, S. Kumar, A. Hanan, J. Hussain, and S. Abdullah, "Drawbacks of Artificial Intelligence and Their Potential Solutions in the Healthcare Sector," Biomedical Materials & Devices (New York, N.y.), pp. 1–8, 2023.
11. S. Jayakumar, V. Sounderajah, P. Normahani, L. Harling, S. R. Markar, H. Ashrafian, and A. Darzi, "Quality assessment standards in artificial intelligence diagnostic accuracy systematic reviews: a metaresearch study," npj Digital Medicine, vol. 5, pp. 1–13, 2022. Number: 1 Publisher: Nature Publishing Group.
12. Trocin C, Mikalef P, Papamitsiou Z, Conboy K. Responsible AI for Digital Health: a Synthesis and a Research Agenda. Inf Syst Front. 2023;25:2139–57.
13. S. C. Rivera, X. Liu, S. E. Hughes, H. Dunster, E. Manna, A. K. Denniston, and M. J. Calvert, "Embedding patient-reported outcomes at the heart of artificial intelligence health-care technologies," The Lancet Digital Health, vol. 5, pp. e168–e173, 2023. Publisher: Elsevier.
14. Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: A call for open science. Patterns. 2021;2: 100347.
15. Engstrom T, Tanner S, Lee WR, Forbes C, Walker R, Bradford N, Pole JD. Patient reported outcome measure domains and tools used among adolescents and young adults with cancer: A scoping review. Crit Rev Oncol Hematol. 2023;181: 103867.
16. P. Jayakumar, E. Lin, V. Galea, A. J. Mathew, N. Panda, I. Vetter, and A. B. Haynes, "Digital Phenotyping and Patient-Generated Health Data for Outcome Measurement in Surgical Care: A Scoping Review," Journal

of Personalized Medicine, vol. 10, p. 282, 2020. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

17. S. S. Hoque, S. Ahern, H. E. O'Connell, L. Romero, and R. Ruseckaite, "Comparing patient-reported outcome measures for pain in women with pelvic floor disorders: a scoping review," The Journal of Pain, 2023.

18. Peters MDJ, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc. 2015;13:141–6.

19. Wojcik Z, Dimitrova V, Warrington L, Velikova G, Absolom K. Predicting patient outcomes from self-reported symptom data using Artificial intelligence: a scoping review protocol. OSF: Publisher; 2023.

20. A. C. Tricco, E. Lillie, W. Zarin, K. K. O'Brien, H. Colquhoun, D. Levac, D. Moher, M. D. Peters, T. Horsley, L. Weeks, S. Hempel, E. A. Akl, C. Chang, J. McGowan, L. Stewart, L. Hartling, A. Aldcroft, M. G. Wilson, C. Garritty, S. Lewin, C. M. Godfrey, M. T. Macdonald, E. V. Langlois, K. Soares-Weiser, J. Moriarty, T. Clifford, Tun¸calp, and S. E. Straus, "PRISMA Extension for Scoping Reviews (PRISMAScR): Checklist and Explanation," Annals of Internal Medicine, vol. 169, pp. 467–473, 2018.

21. Kersting C, Kneer M, Barzel A. Patient-relevant outcomes: what are we talking about? A scoping review to improve conceptual clarity. BMC Health Serv Res. 2020;20:596.

22. E. M. Roos, H. P. Roos, L. S. Lohmander, C. Ekdahl, and B. D. Beynnon, "Knee Injury and Osteoarthritis Outcome Score (KOOS)—Development of a Self-Administered Outcome Measure," Journal of Orthopaedic & Sports Physical Therapy, vol. 28, pp. 88–96, 1998. Publisher: Journal of Orthopaedic & Sports Physical Therapy.

23. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A, Flechtner H, Fleishman SB, de Haes JC. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst. 1993;85:365–76.

24. Rabin R, de Charro F. EQ-5D: a measure of health status from the Euro-Qol Group. Ann Med. 2001;33:337–43.

25. Koyyada SP, Singh TP. A multi stage approach to handle class imbalance: An ensemble method. Procedia Computer Science. 2023;218:2666–74.

26. Wang Y, Van Dijk L, Mohamed ASR, Fuller CD, Zhang X, Marai GE, Canahuate G. "Predicting late symptoms of head and neck cancer treatment using LSTM and patient reported outcomes", Proceedings. International Database Engineering and Applications Symposium. 2021;2021:273–9.

27. M. Kalweit, U. A. Walker, A. Finckh, R. Mu¨ller, G. Kalweit, A. Scherer, J. Boedecker, and T. Hu¨gle, "Personalized prediction of disease activity in patients with rheumatoid arthritis using an adaptive deep neural network," PloS One, vol. 16, no. 6, p. e0252289, 2021.

28. R. C. Grant, J. C. He, F. Khan, N. Liu, S. Podolsky, Y. Kaliwal, M. Powis, F. Notta, K. K. W. Chan, M. Ghassemi, S. Gallinger, and M. K. Krzyzanowska, "Machine Learning–Based Early Warning Systems for Acute Care Utilization During Systemic Therapy for Cancer," Journal of the National Comprehensive Cancer Network, vol. 21, pp. 1029–1037.e21, 2023. Publisher: National Comprehensive Cancer Network Section: Journal of the National Comprehensive Cancer Network.

29. Klemt C, Uzosike AC, Esposito JG, Harvey MJ, Yeo I, Subih M, Kwon Y-M. The utility of machine learning algorithms for the prediction of patient-reported outcome measures following primary hip and knee total joint arthroplasty. Arch Orthop Trauma Surg. 2023;143:2235–45.

30. Tennenhouse LG, Marrie RA, Bernstein CN, Lix LM. Machine-learning models for depression and anxiety in individuals with immune-mediated inflammatory disease. J Psychosom Res. 2020;134: 110126.

31. Zhang S, Chen JY, Pang HN, Lo NN, Yeo SJ, Liow MHL. Development and internal validation of machine learning algorithms to predict patient satisfaction after total hip arthroplasty. Arthroplasty. 2021;3:33.

32. Zhang S, Lau BPH, Ng YH, Wang X, Chua W. Machine learning algorithms do not outperform preoperative thresholds in predicting clinically meaningful improvements after total knee arthroplasty. Knee Surg Sports Traumatol Arthrosc. 2022;30:2624–30.

33. D. Verma, K. Bach, and P. J. Mork, "Using Automated Feature Selection for Building Case-Based Reasoning Systems: An Example from Patient-Reported Outcome Measurements," in Artificial Intelligence XXXVIII: 41st SGAI International Conference on Artificial Intelligence, AI 2021, Cambridge, UK, December 14–16, 2021, Proceedings, (Berlin, Heidelberg), pp. 282–295, Springer-Verlag, 2021.

34. Buus AA, Udsen FW, Laugesen B, El-Galaly A, Laursen M, Hejlesen OK. Patient-Reported Outcomes for Function and Pain in Total Knee Arthroplasty Patients. Nurs Res. 2022;71:E39–47.

35. V. E. Staartjes, M. P. de Wispelaere, W. P. Vandertop, and M. L. Schr¨oder, "Deep learning-based preoperative predictive analytics for patient-reported outcomes following lumbar discectomy: feasibility of center-specific modeling," The Spine Journal: Official Journal of the North American Spine Society, vol. 19, pp. 853–861, 2019.

36. 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.

37. C. Xu, I. M. Subbiah, S.-C. Lu, A. Pfob, and C. Sidey-Gibbons, "Machine learning models for 180-day mortality prediction of patients with advanced cancer using patient-reported symptom data," Quality of Life Research, 2022.

38. A. H. S. Harris, A. C. Kuo, T. R. Bowe, L. Manfredi, N. F. Lalani, and N. J. Giori, "Can Machine Learning Methods Produce Accurate and Easy-to-Use Preoperative Prediction Models of One-Year Improvements in Pain and Functioning After Knee Arthroplasty?," The Journal of Arthroplasty, vol. 36, pp. 112–117.e6, 2021. Publisher: Elsevier.

39. A. Siccoli, M. P. d. Wispelaere, M. L. Schr¨oder, and V. E. Staartjes, "Machine learning–based preoperative predictive analytics for lumbar spinal stenosis," Neurosurgical Focus, vol. 46, p. E5, 2019. Publisher: American Association of Neurological Surgeons Section: Neurosurgical Focus.

40. A. Pfob, B. J. Mehrara, J. A. Nelson, E. G. Wilkins, A. L. Pusic, and C. Sidey-Gibbons, "Machine learning to predict individual patient-reported outcomes at 2-year follow-up for women undergoing cancer-related mastectomy and breast reconstruction (INSPiRED-001)," The Breast, vol. 60, pp. 111–122, Dec. 2021. Publisher: Elsevier.

41. Pedersen CF, Andersen M, Carreon LY, Eiskjær S. Applied Machine Learning for Spine Surgeons: Predicting Outcome for Patients Undergoing Treatment for Lumbar Disc Herniation Using PRO Data. Global Spine Journal. 2022;12:866–76.

42. A. Katakam, A. V. Karhade, A. Collins, D. Shin, C. Bragdon, A. F. Chen, C. M. Melnic, J. H. Schwab, and H. S. Bedair, "Development of machine learning algorithms to predict achievement of minimal clinically important difference for the KOOS-PS following total knee arthroplasty," Journal of Orthopaedic Research, vol. 40, no. 4, pp. 808–815, 2022. eprint: https://onlinelibrary.wiley.com/doi/pdf/https://doi.org/10.1002/jor.25125.

43. C. B. Josephson, J. D. T. Engbers, T. T. Sajobi, S. Wahby, O. A. Lawal, M. R. Keezer, D. K. Nguyen, K. Malmgren, M. J. Atkinson, W. J. Hader, S. Macrodimitris, S. B. Patten, N. Pillay, R. Sharma, S. Singh, Y. Starreveld, and S. Wiebe, "Predicting postoperative epilepsy surgery satisfaction in adults using the 19-item Epilepsy Surgery Satisfaction Questionnaire and machine learning," Epilepsia, vol. 62, no. 9, pp. 2103–2112, 2021. eprint: https://onlinelibrary.wiley.com/doi/pdf/https://doi.org/10.1111/epi.16992.

44. P. N. Ramkumar, J. M. Karnuta, H. S. Haeberle, S. A. Rodeo, B. U. Nwachukwu, and R. J. Williams, "Effect of Preoperative Imaging and Patient Factors on Clinically Meaningful Outcomes and Quality of Life After Osteochondral Allograft Transplantation: A Machine Learning Analysis of Cartilage Defects of the Knee," The American Journal of Sports Medicine, vol. 49, pp. 2177–2186, 2021. Publisher: SAGE Publications Inc STM.

45. J. S. Munn, B. A. Lanting, S. J. MacDonald, L. E. Somerville, J. D. Marsh, D. M. Bryant, and B. M. Chesworth, "Logistic Regression and Machine Learning Models Cannot Discriminate Between Satisfied and Dissatisfied Total Knee Arthroplasty Patients," The Journal of Arthroplasty, vol. 37, pp. 267–273, 2022. Publisher: Elsevier.

46. Kober KM, Roy R, Dhruva A, Conley YP, Chan RJ, Cooper B, Olshen A, Miaskowski C. Prediction of evening fatigue severity in outpatients receiving chemotherapy: less may be more. Fatigue : biomedicine, health & behavior. 2021;9(1):14–32.

47. Durand WM, Daniels AH, Hamilton DK, Passias P, Kim HJ, Protopsaltis T, LaFage V, Smith JS, Shaffrey C, Gupta M, Klineberg E, Schwab F, Burton D, Bess S, Ames C, Hart R. Artificial Intelligence Models Predict Operative Versus Nonoperative Management of Patients with Adult Spinal Deformity with 86% Accuracy. World Neurosurgery. 2020;141:e239–53.

48. C. J. Harrison, L. Geoghegan, C. J. Sidey-Gibbons, P. H. C. Stirling, J. E. McEachan, and J. N. Rodrigues, "Developing Machine Learning Algorithms to Support Patient-centered, Value-based Carpal Tunnel

Decompression Surgery," Plastic and Reconstructive Surgery Global Open, vol. 10, p. e4279, Apr. 2022. V. Kumar, C. Roche, S. Overman, R. Simovitch, P.-H. Flurin, T. Wright, J. Zuckerman, H. Routman, and Teredesai, "What Is the Accuracy of Three Different Machine Learning Techniques to Predict ClinicalOutcomes After Shoulder Arthroplasty?," Clinical Orthopaedics and Related Research®, vol. 478, p. 2351, 2020.

49. Wshah S, Skalka C, Price M. Predicting Posttraumatic Stress Disorder Risk: A Machine Learning Approach", JMIR Mental Health, vol. 6, p. e13946, 2019. Y. Lu, E. Forlenza, R. R. Wilbur, O. Lavoie-Gagne, M. C. Fu, A. B. Yanke, B. J. Cole, N. Verma, and Forsythe, "Machine-learning model successfully predicts patients at risk for prolonged postoperativeopioid use following elective knee arthroscopy. Knee surgery, sports trauma-tology, arthroscopy: official journal of the ESSKA. 2022;30:762–72.

50. Kunze KN, Karhade AV, Sadauskas AJ, Schwab JH, Levine BR. Develop-ment of Machine Learning Algorithms to Predict Clinically Meaningful Improvement for the Patient-Reported Health State After Total Hip Arthroplasty. J Arthroplasty. 2020;35:2119–23.

51. N. L. Loos, L. Hoogendam, J. S. Souer, H. P. Slijper, E.-R. Andrinopou-lou, M. W. Coppieters, R. W. Selles, and , the Hand-Wrist Study Group, "Machine Learning Can be Used to Predict Function but Not Pain After Surgery for Thumb Carpometacarpal Osteoarthritis," Clinical Orthopae-dics and Related Research, vol. 480, pp. 1271–1284, 2022.

52. Shipston-Sharman O, Popkirov S, Hansen CH, Stone J, Carson A. Prog-nosis in functional and recognised pathophysiological neurological disorders - a shared basis. J Psychosom Res. 2021;152: 110681.

53. B. K. Tan, G. Lu, M. J. Kwasny, W. D. Hsueh, S. Shintani-Smith, D. B. Conley, R. K. Chandra, R. C. Kern, and R. Leung, "Effect of symptom-based risk stratification on the costs of managing patients with chronic rhinosi-nusitis symptoms," International Forum of Allergy & Rhinology, vol. 3, no. 11, pp. 933–940, 2013. eprint: https://onlinelibrary.wiley.com/doi/pdf/https://doi.org/10.1002/alr.21208.

54. A. A. Rahman, M. I. Siraji, L. I. Khalid, F. Faisal, M. M. Nishat, A. Ahmed, and M. A. A. Mamun, "Perceived Stress Analysis of Undergraduate Students during COVID-19: A Machine Learning Approach," in 2022 IEEE 21st Mediterranean Electrotechnical Conference (MELECON), pp. 1129–1134, 2022. ISSN: 2158–8481.

55. E. S. Lee, "Exploring the Performance of Stacking Classifier to Predict Depression Among the Elderly," in 2017 IEEE International Conference on Healthcare Informatics (ICHI), pp. 13–20, 2017.

56. Nowinka Z, Alagha MA, Mahmoud K, Jones GG. Predicting Depression in Patients With Knee Osteoarthritis Using Machine Learning: Model Development and Validation Study. JMIR formative research. 2022;6: e36130.

57. Arkin FS, Aras G, Dogu E. Comparison of Artificial Neural Networks and Logistic Regression for 30-days Survival Prediction of Cancer Patients. Acta informatica medica: AIM: journal of the Society for Medical Informatics of Bosnia & Herzegovina: casopis Drustva za medicinsku informatiku BiH. 2020;28:108–13.

58. L. Hoogendam, J. A. C. Bakx, J. S. Souer, H. P. Slijper, E.-R. Andrinopoulou, R. W. Selles, and Hand Wrist Study Group, "Predicting Clinically Relevant Patient-Reported Symptom Improvement After Carpal Tunnel Release: A Machine Learning Approach," Neurosurgery, vol. 90, pp. 106–113, 2022.

59. Camp EJ, Quon RJ, Sajatovic M, Briggs F, Brownrigg B, Janevic MR, Meisenhelter S, Steimel SA, Testorf ME, Kiriakopoulos E, Mazanec MT, Fraser RT, Johnson EK, Jobst BC. "Supervised machine learning to predict reduced depression severity in people with epilepsy through epilepsy self-management intervention", Epilepsy & Behavior, vol. 127. Elsevier: Publisher; 2022.

60. Wardenaar KJ, Riese H, Giltay EJ, Eikelenboom M, van Hemert AJ, Beekman AF, Penninx BWJH, Schoevers RA. Common and specific determinants of 9-year depression and anxiety course-trajectories: A machine-learning investigation in the Netherlands Study of Depression and Anxiety (NESDA). J Affect Disord. 2021;293:295–304.

61. J.-a. Sim, Y. A. Kim, J. H. Kim, J. M. Lee, M. S. Kim, Y. M. Shim, J. I. Zo, and Y. H. Yun, "The major effects of health-related quality of life on 5-year sur-vival prediction among lung cancer survivors: applications of machine learning," Scientific Reports, vol. 10, p. 10693, 2020. Number: 1 Publisher: Nature Publishing Group.

62. K. N. Kunze, E. M. Polce, B. U. Nwachukwu, J. Chahla, and S. J. Nho, "Development and Internal Validation of Supervised Machine Learning

Algorithms for Predicting Clinically Significant Functional Improvement in a Mixed Population of Primary Hip Arthroscopy," Arthroscopy, vol. 37, pp. 1488–1497, 2021. Publisher: Elsevier.

63. Smith DL, Held P. Moving toward precision PTSD treatment: predicting veterans' intensive PTSD treatment response using continuously updat-ing machine learning models. Psychol Med. 2023;53:5500–9.

64. Bone C, Simmonds-Buckley M, Thwaites R, Sandford D, Merzhvynska M, Rubel J, Deisenhofer A-K, Lutz W, Delgadillo J. Dynamic prediction of psychological treatment outcomes: development and validation of a prediction model using routinely collected symptom data. The Lancet Digital Health. 2021;3:e231–40.

65. O'Driscoll C, Buckman JEJ, Fried EI, Saunders R, Cohen ZD, Ambler G, DeRubeis RJ, Gilbody S, Hollon SD, Kendrick T, Kessler D, Lewis G, Watkins E, Wiles N, Pilling S. The importance of transdiagnostic symptom level assessment to understanding prognosis for depressed adults: analysis of data from six randomised control trials. BMC Med. 2021;19:109.

66. Shi H-Y, Tsai J-T, Chen Y-M, Culbertson R, Chang H-T, Hou M-F. Predict-ing two-year quality of life after breast cancer surgery using artificial neural network and linear regression models. Breast Cancer Res Treat. 2012;135:221–9.

67. Pua Y-H, Kang H, Thumboo J, Clark RA, Chew ES-X, Poon CL-L, Chong H-C, Yeo S-J. Machine learning methods are comparable to logistic regression techniques in predicting severe walking limitation fol-lowing total knee arthroplasty. Knee Surg Sports Traumatol Arthrosc. 2020;28:3207–16.

68. Langenberger B. Who will stay a little longer? Predicting length of stay in hip and knee arthroplasty patients using machine learning. Intelligence-Based Medicine. 2023;8: 100111.

69. Hasannejadasl H, Osong B, Bermejo I, van der Poel H, Vanneste B, van Roermund J, Aben K, Zhang Z, Kiemeney L, Van Oort I, Verwey R, Hoch-stenbach L, Bloemen E, Dekker A, Fijten RRR. A comparison of machine learning models for predicting urinary incontinence in men with local-ized prostate cancer. Front Oncol. 2023;13:1168219.

70. Xu C, Pfob A, Mehrara BJ, Yin P, Nelson JA, Pusic AL, Sidey-Gibbons C. Enhanced Surgical Decision-Making Tools in Breast Cancer: Predict-ing 2-Year Postoperative Physical, Sexual, and Psychosocial Well-Being following Mastectomy and Breast Reconstruction (INSPiRED 004). Ann Surg Oncol. 2023;30:7046–59.

71. Tian J, Yan J, Han G, Du Y, Hu X, He Z, Han Q, Zhang Y. Machine learning prognosis model based on patient-reported outcomes for chronic heart failure patients after discharge. Health Qual Life Outcomes. 2023;21:31.

72. C. Park, P. V. Mummaneni, O. N. Gottfried, C. I. Shaffrey, A. J. Tang, E. F. Bisson, A. L. Asher, D. Coric, E. A. Potts, K. T. Foley, M. Y. Wang, K.-M. Fu, M. S. Virk, J. J. Knightly, S. Meyer, P. Park, Upadhyaya, M. E. Shaffrey, A. L. Buchholz, L. M. Tumial´an, J. D. Turner, B. A. Sherrod, N. Agarwal,D. Chou, R. W. Haid, M. Bydon, and A. K. Chan, "Which supervised machine learning algorithm can best predict achievement of minimum clinically important difference in neck pain after surgery in patients with cervical myelopathy? A QOD study," Neurosurgical Focus, vol. 54, p. E5, 2023. Publisher: American Association of Neurological Surgeons Section: Neurosurgical Focus.

73. Langenberger B, Schrednitzki D, Halder AM, Busse R, Pross CM. Predicting whether patients will achieve minimal clinically important differences following hip or knee arthroplasty. Bone & Joint Research. 2023;12:512–21.

74. Curtis JR, Su Y, Black S, Xu S, Langholff W, Bingham CO, Kafka S, Xie F. Machine Learning Applied to Patient-Reported Outcomes to Classify Physician-Derived Measures of Rheumatoid Arthritis Disease Activity. ACR open rheumatology. 2022;4:995–1003.

75. Iivanainen S, Ekstrom J, Virtanen H, Kataja VV, Koivunen JP. Electronic patient-reported outcomes and machine learning in predicting immune-related adverse events of immune checkpoint inhibitor thera-pies. BMC Med Inform Decis Mak. 2021;21:205.

76. Pfob A, Mehrara BJ, Nelson JA, Wilkins EG, Pusic AL, Sidey-Gibbons C. Towards Patient-centered Decision-making in Breast Cancer Surgery: Machine Learning to Predict Individual Patient-reported Outcomes at 1-year Follow-up. Ann Surg. 2023;277: e144.

77. C. W. Noel, R. Sutradhar, L. Gotlib Conn, D. Forner, W. C. Chan, R. Fu, J. Hallet, N. G. Coburn, and A. Eskander, "Development and Validation of

Wójcik *et al. Health and Quality of Life Outcomes*     (2025) 23:37

Page 14 of 15

a Machine Learning Algorithm Predicting Emergency Department Use and Unplanned Hospitalization in Patients With Head and Neck Cancer," JAMA otolaryngology– head & neck surgery, vol. 148, pp. 764–772, 2022.

78. Chmiel FP, Burns DK, Pickering JB, Blythin A, Wilkinson TM, Boniface MJ. Prediction of Chronic Obstructive Pulmonary Disease Exacerbation Events by Using Patient Self-reported Data in a Digital Health App: Statistical Evaluation and Machine Learning Approach. JMIR Med Inform. 2022;10: e26499.

79. H. N. Ziobrowski, C. J. Kennedy, B. Ustun, S. L. House, F. L. Beaudoin, X. An, D. Zeng, K. A. Bollen, M. Petukhova, N. A. Sampson, V. Puac-Polanco, S. Lee, K. C. Koenen, K. J. Ressler, S. A. McLean, R. C. Kessler, AURORA Consortium, J. S. Stevens, T. C. Neylan, G. D. Clifford, T. Jovanovic, S. D. Linnstaedt, L. T. Germine, S. L. Rauch, J. P. Haran, A. B. Storrow, C. Lewandowski, P. I. Musey, P. L. Hendry, S. Sheikh, C. W. Jones, B. E. Punches, M. S. Lyons, V. P. Murty, M. E. McGrath, J. L. Pascual, M. J. Seamon, E. M. Datner, A. M. Chang, C. Pearson, D. A. Peak, G. Jambaulikar, R. C. Merchant, R. M. Domeier, N. K. Rathlev, B. J. O'Neil, P. Sergot, L. D. Sanchez, S. E. Bruce, R. H. Pietrzak, J. Joormann, D. M. Barch, D. A. Pizzagalli, J. F. Sheridan, S. E. Harte, J. M. Elliott, and S. J. H. van Rooij, "Development and Validation of a Model to Predict Posttraumatic Stress Disorder and Major Depression After a Motor Vehicle Collision," JAMA psychiatry, vol. 78, pp. 1228–1237, 2021.

80. P. Annapureddy, M. F. Hossain, T. Kissane, W. Frydrychowicz, P. Nitu, J. Coelho, N. Johnson, P. Madiraju, Z. Franco, K. Hooyer, N. Jain, M. Flower, and S. Ahamed, "Predicting PTSD Severity in Veterans from Self-reports for Early Intervention: A Machine Learning Approach," in 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI), pp. 201–208, 2020.

81. M. Zhong, V. Van Zoest, A. M. Bilal, F. Papadopoulos, and G. Castellano, "Unimodal vs. Multimodal Prediction of Antenatal Depression from Smartphone-based Survey Data in a Longitudinal Study," in Proceedings of the 2022 International Conference on Multimodal Interaction, ICMI '22, (New York, NY, USA), pp. 455–467, Association for Computing Machinery, 2022.

82. C. J. Sidey-Gibbons, C. Sun, A. Schneider, S.-C. Lu, K. Lu, A. Wright, and L. Meyer, "Predicting 180day mortality for women with ovarian cancer using machine learning and patient-reported outcome data," Scientific Reports, vol. 12, p. 21269, 2022. Number: 1 Publisher: Nature Publishing Group.

83. L. S. Canas, C. H. Sudre, J. C. Pujol, L. Polidori, B. Murray, E. Molteni, M. S. Graham, K. Klaser, M. Antonelli, S. Berry, R. Davies, L. H. Nguyen, D. A. Drew, J. Wolf, A. T. Chan, T. Spector, C. J. Steves, S. Ourselin, and M. Modat, "Early detection of COVID-19 in the UK using self-reported symptoms: a large-scale, prospective, epidemiological surveillance study," The Lancet Digital Health, vol. 3, pp. e587– e598, 2021. Publisher: Elsevier.

84. R. Sutradhar and L. Barbera, "Comparing an Artificial Neural Network to Logistic Regression for Predicting ED Visit Risk Among Patients With Cancer: A Population-Based Cohort Study," Journal of Pain and Symptom Management, vol. 60, pp. 1–9, 2020. Publisher: Elsevier.

85. Goldstein N, Eisenkraft A, Arguello CJ, Yang GJ, Sand E, Ishay AB, Merin R, Fons M, Littman R, Nachman D, Gepner Y. Exploring Early Pre-Symptomatic Detection of Influenza Using Continuous Monitoring of Advanced Physiological Parameters during a Randomized Controlled Trial. J Clin Med. 2021;10:5202.

86. Rahman QA, 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.

87. Verma D, Jansen D, Bach K, Poel M, Mork PJ, d'Hollosy WON. Exploratory application of machine learning methods on patient reported data in the development of supervised models for predicting outcomes. BMC Med Inform Decis Mak. 2022;22:227.

88. Crowson MG, Dixon P, Mahmood R, Lee JW, Shipp D, Le T, Lin V, Chen J, Chan TCY. Predicting Postoperative Cochlear Implant Performance Using Supervised Machine Learning. Otol Neurotol. 2020;41: e1013.

89. Milella F, Famiglini L, Banfi G, Cabitza F. Application of Machine Learning to Improve Appropriateness of Treatment in an Orthopaedic Setting of Personalized Medicine. Journal of Personalized Medicine. 2022;12:1706.

90. E. M. Polce, K. N. Kunze, M. C. Fu, G. E. Garrigues, B. Forsythe, G. P. Nicholson, B. J. Cole, and N. N. Verma, "Development of supervised machine learning algorithms for prediction of satisfaction at 2 years following total shoulder arthroplasty," Journal of Shoulder and Elbow Surgery, vol. 30, pp. e290–e299, 2021. Publisher: Elsevier.

91. P. Thanathamathee, "Boosting with feature selection technique for screening and predicting adolescents depression," in 2014 Fourth International Conference on Digital Information and Communication Technology and its Applications (DICTAP), pp. 23–27, 2014.

92. J. A. A. Mendoza, G. A. Solano, M. J. Pontiveros, J. Dl Caro, P. M. D. Gomez, C. G. Manuel, P. J. B. Rosell-Ubial, and M. Tee, "A Machine Learning Approach in Evaluating Symptom Screening in Predicting COVID-19," in 2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), pp. 188–193, 2022.

93. Z. Xu, Y. Wang, and B. Jiang, "Research on Prisoner Psychological Symptoms Quick Screening Model Based on Ensemble Learning," in Proceedings of the 3rd International Symposium on Artificial Intelligence for Medicine Sciences, ISAIMS '22, (New York, NY, USA), pp. 563–567, Association for Computing Machinery, Dec. 2022.

94. Fu MR, Wang Y, Li C, Qiu Z, Axelrod D, Guth AA, Scagliola J, Conley Y, Aouizerat BE, Qiu JM, Yu G, Cleave JHV, Haber J, Cheung YK. "Machine learning for detection of lymphedema among breast cancer survivors", mHealth, vol. 4. AME publishing company: Publisher; 2018.

95. X. Pan, R. Levin-Epstein, J. Huang, D. Ruan, C. R. King, A. U. Kishan, M. L. Steinberg, and X. S. Qi, "Dosimetric predictors of patient-reported toxicity after prostate stereotactic body radiotherapy: Analysis of full range of the dose–volume histogram using ensemble machine learning," Radiotherapy and Oncology, vol. 148, pp. 181–188, 2020. Publisher: Elsevier.

96. Agochukwu-Mmonu N, Murali A, Wittmann D, Denton B, Dunn RL, Montie J, Peabody J, Miller D, Singh K. Development and Validation of Dynamic Multivariate Prediction Models of Sexual Function Recovery in Patients with Prostate Cancer Undergoing Radical Prostatectomy: Results from the MUSIC Statewide Collaborative. European Urology Open Science. 2022;40:1–8.

97. A. M. Chekroud, R. J. Zotti, Z. Shehzad, R. Gueorguieva, M. K. Johnson, M. H. Trivedi, T. D. Cannon, J. H. Krystal, and P. R. Corlett, "Cross-trial prediction of treatment outcome in depression: a machine learning approach," The Lancet Psychiatry, vol. 3, pp. 243–250, 2016. Publisher: Elsevier.

98. W. Oude Nijeweme-d'Hollosy, L. van Velsen, M. Poel, C. G. M. Groothuis-Oudshoorn, R. Soer, and H. Hermens, "Evaluation of three machine learning models for self-referral decision support on low back pain in primary care," International Journal of Medical Informatics, vol. 110, pp. 31–41, 2018.

99. A. Goldstein and S. Cohen, "Self-report symptom-based endometriosis prediction using machine learning," Scientific Reports, vol. 13, p. 5499, 2023. Number: 1 Publisher: Nature Publishing Group.

100. S. Iivanainen, J. Ekstr¨om, H. Virtanen, V. V. Kataja, and J. P. Koivunen, "Predicting Objective Response Rate (ORR) in Immune Checkpoint Inhibitor (ICI) Therapies with Machine Learning (ML) by Combining Clinical and Patient-Reported Data," Applied Sciences, vol. 12, p. 1563, 2022. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

101. G. Luo, B.L. Stone, B. Fassl, et al. "Predicting asthma control deterioration in children. "BMC Med Inform Decis Mak 15, 84, 2015)

102. Verma D, Bach K, Mork PJ. External validation of prediction models for patient-reported outcome measurements collected using the selfBACK mobile app. Int J Med Informatics. 2023;170: 104936.

103. Haeberle HS, Ramkumar PN, Karnuta JM, Sullivan S, Sink EL, Kelly BT, Ranawat AS, Nwachukwu BU. Predicting the risk of subsequent hip surgery before primary hip arthroscopy for femoroacetabular impingement syndrome: a machine learning analysis of preoperative risk factors in hip preservation. Am J Sports Med. 2021;49:2668–76.

104. Sun R, Tomkins-Lane C, Muaremi A, Kuwabara A, Smuck M. Physical activity thresholds for predicting longitudinal gait decline in adults with knee osteoarthritis. Osteoarthritis Cartilage. 2021;29:965–72.

105. K. Schultebraucks, M. Qian, D. Abu-Amara, K. Dean, E. Laska, C. Siegel, A. Gautam, G. Guffanti, R. Hammamieh, B. Misganaw, S. H. Mellon, O. M. Wolkowitz, E. M. Blessing, A. Etkin, K. J. Ressler, F. J. Doyle, M. Jett, and C. R. Marmar, "Pre-deployment risk factors for PTSD in active-duty personnel deployed to Afghanistan: a machine-learning approach

Wójcik *et al. Health and Quality of Life Outcomes*    (2025) 23:37

Page 15 of 15

for analyzing multivariate predictors," Molecular Psychiatry, vol. 26, pp. 5011–5022, 2021. Number: 9 Publisher: Nature Publishing Group.

106. Tschuggnall M, Grote V, Pirchl M, Holzner B, Rumpold G, Fischer MJ. Machine learning approaches to predict rehabilitation success based on clinical and patient-reported outcome measures. Informatics in Medicine Unlocked. 2021;24: 100598.

107. M. H. Sandham, E. A. Hedgecock, R. J. Siegert, A. Narayanan, M. B. Hocaoglu, and I. J. Higginson, "Intelligent Palliative Care Based on Patient-Reported Outcome Measures," Journal of Pain and Symptom Management, vol. 63, pp. 747–757, 2022. Publisher: Elsevier.

108. H. Pappot, B. P. Bj¨ornsson, O. Krause, C. Bæksted, P. E. Bidstrup, S. O. Dalton, C. Johansen, A. Knoop, I. Vogelius, and C. Holl¨ander-Mieritz, "Machine learning applied in patient-reported outcome research— exploring symptoms in adjuvant treatment of breast cancer," Breast Cancer, vol. 31, pp. 148–153, 2024.

109. Y.-C. Hsu, J.-D. Wang, P.-H. Huang, Y.-W. Chien, C.-J. Chiu, and C.-Y. Lin,"Integrating domain knowledge with machine learning to detect obstructive sleep apnea:Snore as a significant bio-feature,"Journal of Sleep Research, vol. 31, no. 2, p. e13487, 2022. eprint: https://onlineli-brary.wiley.com/doi/pdf/https://doi.org/10.1111/jsr.13487.

110. I. Miranda, G. Cardoso, M. Pahar, G. Oliveira, and T. Niesler, "Machine Learning Prediction of Hospitalization due to COVID-19 based on Self-Reported Symptoms: A Study for Brazil," in 2021 IEEEEMBS International Conference on Biomedical and Health Informatics (BHI), pp. 1–5, 2021. ISSN: 2641–3604.

111. Long MJ, Papi E, Duffell LD, McGregor AH. Predicting knee osteoarthritis risk in injured populations. Clin Biomech. 2017;47:87–95.

112. Xuyi W, Seow H, Sutradhar R. Artificial neural networks for simultane-ously predicting the risk of multiple co-occurring symptoms among patients with cancer. Cancer Med. 2020;10:989–98.

113. Shahzad MN, Suleman M, Ahmed MA, Riaz A, Fatima K. Identifying the Symptom Severity in Obsessive-Compulsive Disorder for Classification and Prediction: An Artificial Neural Network Approach. Behav Neurol. 2020;2020:2678718.

114. A. Bugajski, A. Lengerich, R. Koerner, and L. Szalacha, "Utilizing an Artificial Neural Network to Predict Self-Management in Patients With Chronic Obstructive Pulmonary Disease: An Exploratory Analysis," Jour-nal of Nursing Scholarship, vol. 53, no. 1, pp. 16–24, 2021. eprint: https://onlinelibrary.wiley.com/doi/pdf/https://doi.org/10.1111/jnu.12618.

115. L. van der Stap, M. F. van Haaften, E. F. van Marrewijk, A. H. de Heij, P. L. Jansen, J. M. N. Burgers, M. S. Sieswerda, R. K. Los, A. K. L. Reyners, and Y. M. van der Linden, "The feasibility of a Bayesian network model to assess the probability of simultaneous symptoms in patients with advanced cancer," Scientific Reports, vol. 12, p. 22295, 2022. Number: 1 Publisher: Nature Publishing Group.

116. T. Strating, L. Shafiee Hanjani, I. Tornvall, R. Hubbard, and I. A. Scott, "Navigating the machine learning pipeline: a scoping review of inpa-tient delirium prediction models," BMJ Health & Care Informatics, vol. 30, p. e100767, 2023.

117. S. Cruz Rivera, X. Liu, S. E. Hughes, H. Dunster, E. Manna, A. K. Denniston, and M. J. Calvert, "Embedding patient-reported outcomes at the heart of artificial intelligence health-care technologies,"The Lancet Digital Health, vol. 5, pp. e168–e173, 2023.

118. Whitebird RR, Solberg LI, Ziegenfuss JY, Norton CK, Chrenka EA, Swiontkowski M, Reams M, Grossman ES. What Do Orthopaedists Believe is Needed for Incorporating Patient-reported Outcome Measures into Clinical Care? A Qualitative Study. Clin Orthop Relat Res. 2022;480:680–7.

119. M. R. Davies, J. E. J. Buckman, B. N. Adey, C. Armour, J. R. Bradley, S. C. B. Curzons, H. L. Davies, K. A. S. Davis, K. A. Goldsmith, C. R. Hirsch, M. Hotopf, C. Hu¨bel, I. R. Jones, G. Kalsi, G. Krebs, Y. Lin, I. Marsh, M. McAtarsney-Kovacs, A. M. McIntosh, J. Mundy, D. Monssen, A. J. Peel, H. C. Rogers, M. Skelton, D. J. Smith, A. ter Kuile, K. N. Thompson, D. Veale, J. T. R. Walters, R. Zahn, G. Breen, and T. C. Eley, "Comparison of symptom-based versus self-reported diagnostic measures of anxiety and depression disorders in the GLAD and COPING cohorts," Journal of Anxiety Disorders, vol. 85, p. 102491, 2022.

120. S. Schmeelk, A. Davis, Q. Li, C. Shippey, M. Utah, A. Myers, M. R. Turchioe, and R. M. Creber, "Monitoring Symptoms of COVID-19: Review of Mobile Apps," JMIR mHealth and uHealth, vol. 10, p. e36065, June 2022. Company: JMIR mHealth and uHealth Distributor: JMIR mHealth and

uHealth Institution: JMIR mHealth and uHealth Label: JMIR mHealth and uHealth Publisher: JMIR Publications Inc., Toronto, Canada.

121. Sharma M, Savage C, Nair M, Larsson I, Svedberg P, Nygren JM. Artificial Intelligence Applications in Health Care Practice: Scoping Review. J Med Internet Res. 2022;24: e40238.

122. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: A gentle introduction to imputation of missing values. J Clin Epidemiol. 2006;59:1087–91.

123. D. M. P. Murti, U. Pujianto, A. P. Wibawa, and M. I. Akbar, "K-Nearest Neighbor (K-NN) based Missing Data Imputation," in 2019 5th Interna-tional Conference on Science in Information Technology (ICSITech), pp. 83–88, 2019.

124. Y. Liu, Y. Li, and D. Xie, "Implications of imbalanced datasets for empirical ROC-AUC estimation in binary classification tasks," Journal of Statistical Computation and Simulation, vol. 94, pp. 183–203, 2024. Publisher: Taylor & Francis eprint: https://doi.org/10.1080/00949655.2023.2238235.

125. W. Zuzanna, Dimitrova Vania, Warrington Lorraine, Velikova Galina, and Absolom Kate, "Using Machine Learning to Predict Unplanned Hospital Utilization and Chemotherapy Management From PatientReported Outcome Measures," JCO Clinical Cancer Informatics (In press), 2024.

126. D. Jha, A. Rauniyar, A. Srivastava, D. H. Hagos, N. K. Tomar, V. Sharma, E. Keles, Z. Zhang, U. Demir, A. Topcu, A. Yazidi, J. E. H˚aakeg˚ard, and U. Bagci, "Ensuring Trustworthy Medical Artificial Intelligence through Ethi-cal and Philosophical Principles," 2023. arXiv:2304.11530 [cs].

127. Nisar D-E-M, Amin R, Shah N-U-H, Ghamdi MAA, Almotiri SH, Alruily M. Healthcare Techniques Through Deep Learning: Issues, Challenges and Opportunities. IEEE Access. 2021;9:98523–41.

128. Morid MA, Sheng ORL, Dunbar J. Time Series Prediction Using Deep Learning Methods in Healthcare. ACM Trans Manag Inf Syst. 2023;14:1–29.

129. I. R. Ward, L. Wang, J. lu, M. Bennamoun, G. Dwivedi, and F. M. Sanfilippo, "Explainable Artificial Intelligence for Pharmacovigilance: What Features Are Important When Predicting Adverse Outcomes?,"Computer Meth-ods and Programs in Biomedicine, vol. 212, p. 106415, 2021. arXiv:2112.13210 [cs, q-bio].

130. Kelly BS, Kirwan A, Quinn MS, Kelly AM, Mathur P, Lawlor A, Killeen RP. The ethical matrix as a method for involving people living with disease and the wider public (PPI) in near-term artificial intelligence research. Radiography. 2023;29:S103–11.

131. Y. Chen, A. A. Hosin, M. J. George, F. W. Asselbergs, and A. D. Shah, "Digi-tal technology and patient and public involvement (PPI) in routine care and clinical research—A pilot study," PLOS ONE, vol. 18, p. e0278260, 2023. Publisher: Public Library of Science.

132. W. Xuyi, H. Seow, and R. Sutradhar, "Artificial neural networks for simulta-neously predicting the risk of multiple co-occurring symptoms among patients with cancer," Cancer Medicine, vol. 10, no. 3, pp. 989– 998, 2021. eprint: https://onlinelibrary.wiley.com/doi/pdf/https://doi.org/10.1002/cam4.3685.

133. L.-m. Wang, J.-x. Chen, Y. Pei, X. Zhao, H.-t. Cui, and H.-z. Cui, "Feature Selection and Prediction of Sub-health State Using SVM-RFE," in 2010 International Conference on Artificial Intelligence and Computational Intelligence, vol. 3, pp. 199–202, 2010.

## Publisher's Note