Review

Evidence for Digital Health Tools Designed to Support the Triage of Musculoskeletal Conditions in Primary, Urgent, and Emergency Care Settings: Scoping Review

Linda K Truong[1,2], PT, PhD; James G Wrightson[3], PhD; Raphaël Vincent[4,5,6], PT, MSc; Eunice Lui[1], MPH; Jamon L Couch[7,8], PT; Ellen Wang[2,8], MSc; Cobie Starcevich[9,10], PT; Dean Giustini[11], PhD; Alex Haagaard[12], MA; Elena Lopatina[13,14], MD, PhD; Niels van Berkel[15], PhD; Michael Skovdal Rathleff[16,17], PT, PhD; Clare L Ardern[1,2,7], PT, PhD

[1]Centre for Aging SMART, University of British Columbia, Vancouver, BC, Canada

[2]Department of Physical Therapy, Faculty of Medicine, University of British Columbia, Surrey, BC, Canada

[3]Department of Family Practice, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

[4]School of Rehabilitation, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada

[5]Hôpital Maisonneuve-Rosemont Research Center, Université de Montréal Affiliated Research Center, Montreal, QC, Canada

[6]Centre for Interdisciplinary Research in Rehabilitation of Greater Montreal, Institut Universitaire sur la Réadaptation en Déficience Physique de Montréal, Montreal, QC, Canada

[7]La Trobe Sport and Exercise Medicine Research Centre, La Trobe University, Melbourne, Australia

[8]Arthritis Research Canada, Vancouver, BC, Canada

[9]School of Allied Health, Faculty of Health Sciences, Curtin University, Perth, Australia

[10]Physiotherapy Department, Rockingham General Hospital, South Metropolitan Health Service, Perth, Australia

[11]Biomedical Branch Library, University of British Columbia, Vancouver, BC, Canada

[12]Pain BC, Vancouver, BC, Canada

[13]University of Calgary, Calgary, AB, Canada

[14]Primary Care Alberta, Calgary, AB, Canada

[15]Department of Computer Science, Aalborg University, Aalborg, Denmark

[16]Department of Health Science and Technology, Faculty of Medicine, Aalborg University, Aalborg, Denmark

[17]Center for General Practice at Aalborg University, Aalborg, Denmark

Corresponding Author:

Linda K Truong, PT, PhD
Department of Physical Therapy
Faculty of Medicine, University of British Columbia
13737 96 Avenue
Surrey, BC V3V 0C6
Canada
Phone: 1 604 822 4519
Email: linda.truong@ubc.ca

Related Article:

This is a corrected version. See correction statement in: https://www.jmir.org/2026/1/e92722

Abstract

Background: The digital health research field is growing rapidly, and a summary of the available digital tools for triaging musculoskeletal conditions is needed. Effective and safe digital triage tools for musculoskeletal conditions could support patients and clinicians in making informed care decisions and may contribute to reducing emergency department overcrowding and health care costs.

Objective: The aim of the study is to identify and describe digital health tools for use by adults to triage musculoskeletal conditions across primary, urgent, or emergency care settings.

Methods: Our scoping review was conducted following the Johanna Briggs Institute recommendations for scoping reviews and Arksey and O'Malley's framework. Systematic searches in MEDLINE (OVID), Embase (OVID), CENTRAL (OVID),

CINAHL (EBSCO), Compendex, Web of Science, OpenGrey, Google Scholar, arXiv, medRxiv, and an extensive gray literature search were conducted with a librarian scientist from inception to September 18, 2025. Studies had to recruit adults (aged 18 years and older) with musculoskeletal conditions that identified a digital health tool designed to triage or diagnose in primary, urgent, or emergency care settings and report primary data to be included. In total, 2 reviewer pairs independently screened abstracts and full-text papers. Relevant data were extracted in duplicate, and results were summarized descriptively.

Results: The search yielded 5695 records, and we screened 189 full-text papers. In total, 34 studies (n=37,509 patients) met the inclusion criteria. The most common musculoskeletal conditions reported were rheumatoid or inflammatory arthritis (13/34, 38%). In total, 19 (19/34, 56%) studies reported on symptom checkers, 13 (13/34, 38%) studies on triage or diagnosis tools, and 2 (2/34, 6%) were studies of diagnostic predictor tools. There were 16 unique digital health tools. A total of 2 tools were built for triaging musculoskeletal conditions and were not publicly available outside the UK National Health Service. Most tools were generic tools designed to screen for general health problems, including musculoskeletal conditions. The most common approach to evaluating performance (eg, accuracy) of the tools was to compare the concordance of the tool to a clinician diagnosis or triage recommendation. Sensitivity and specificity ranged from 39% to 91% and 23% to 80%, respectively. The reported accuracy of the included tools ranged from 33% to 98%.

Conclusions: Musculoskeletal conditions remain a blind spot for people designing, implementing, and evaluating digital health for triage: few tools were specifically designed for musculoskeletal conditions, and most existing tools performed poorly when applied to musculoskeletal populations. We recommend health systems and clinicians use a multimodal approach, integrating both digital health tools and clinical decision-making to safely triage and diagnose until a more robust tool for musculoskeletal conditions is available. Future tool developers need to use transparent, standardized processes that prioritize tool safety, clinical value, and trustworthiness when designing for clinicians and patients.

Introduction

Musculoskeletal conditions are one of the largest contributors to the global burden of disease and the sixth largest contributor to disability worldwide [1]. The global forecast predicts that the burden of musculoskeletal conditions will more than double in the decades between 2020 and 2050 [1]. Most musculoskeletal conditions can be effectively managed proactively in a primary care setting, not in the emergency department (ED) [2]. Our recent analysis of epidemiological data revealed that approximately 1 in 10 ED visits were related to musculoskeletal issues, and 6 in 10 of these cases could have been appropriately managed outside the ED [3]. This indicates a need to re-evaluate how people with musculoskeletal conditions access health care.

Effective and efficient triage processes are needed to help patients navigate health systems and find timely and high-quality care, avoiding inappropriate use of the ED [4]. The idea of triage was first applied in military settings to help allocate resources and timely care for the wounded [4]. In today's context, triage is often considered in the ED or the first point of contact for clinicians to help prioritize who needs attention first when patients present to the ED [4]. In regard to triaging musculoskeletal conditions, triage is often conducted by tele-triage, paper-based triage, and face-to-face triage [5]. However, patients and musculoskeletal experts have reported that these approaches are inefficient and ineffective in moving patients through the health system [5].

There is an increasing trend toward the use of digital triage, such as online symptom checkers by patients, to make an informed decision on the next and best course of action for their current problem [6-8]. More recently, the World Health Organization has launched a global strategy on digital health to help improve the health and well-being of all humans [9]. This includes defining digital health as "the use of information and communications technology in support of health and health-related fields," which encompasses eHealth, mobile health (mHealth), advanced computer sciences, such as big data and artificial intelligence (AI) or machine learning, and the broad scope of telehealth and telemedicine [10]. Digital health tools have the potential to tackle overcrowding in the ED and primary care settings by guiding patients to alternative services for musculoskeletal care that may be just as effective as the ED.

In the last decade, there has been a shift toward integrating digital health tools, such as symptom checkers, into the health system, as reflected in the volume of reviews evaluating such tools [6,8,11-14]. With the proliferation and widespread adoption of generative AI (eg, large language models [LLMs]), the public seems accepting of using digital health tools like AI to provide guidance and diagnoses for health conditions. This is despite generative AI not being specifically designed for health care use [15]. These findings reflect the growing demand for the integration of digital technology into health care.

Despite the advancement of digital health tools, there are no reviews currently available that have studied digital health tools for diagnosing and triaging musculoskeletal conditions [8,11-14]. Understanding the available tools, including their performance (eg, accuracy), will help researchers, policymakers, and clinicians tailor future digital health technologies for musculoskeletal conditions and make informed decisions about how to implement technology in health systems. Helping patients find the "right care at the right time" for

musculoskeletal conditions may help reduce burden on the health system and allow the ED to do what it was created for: provide life-saving care.

The primary objective of this review was to identify and describe available digital health tools that can triage and diagnose musculoskeletal conditions in primary, urgent, and emergency settings. The secondary objective was to summarize the performance and accuracy of digital health tools.

Methods

Overview

This scoping review was conducted in accordance with the Johanna Briggs Institute methodology for scoping reviews [16,17] and reported following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews; Checklist 1) and PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Search) [18,19]. We were guided by Arksey and O'Malley's [20] framework with the additions of Levac et al [21]. A protocol was prospectively registered on the Open Science Framework (OSF). Amendments to the protocol were updated and uploaded to the OSF [22].

Search Strategy

An electronic search was conducted in 6 databases (MED-LINE [OVID], Embase [OVID], CENTRAL [OVID], CINAHL [EBSCO], Compendex, and Web of Science) and 4 gray literature sites (OpenGrey, GoogleScholar, arXiv, and medRxiv) with the aid of a biomedical librarian and information specialist. Our search strategy was not peer-reviewed but was tested through an iterative process by the biomedical librarian to ensure that search strategies returned identified seed papers. Initial search strategies were adapted from previously published work [11,14]. We searched databases and gray literature from inception to September 18, 2025. Table 1 provides the population-concept-context framework for our search strategy and illustrates how we operationalized our search. The full MEDLINE search strategy is outlined in Multimedia Appendix 1, and all other search strategies are uploaded and available on OSF [22]. To supplement the search, we screened the reference lists of relevant reviews and included records. We also searched the Cochrane Database of Systematic Reviews, PROSPERO, OSF, and *JBI Evidence Synthesis* to identify any active systematic or scoping reviews on the topic. The search approach for identifying gray literature is detailed in Multimedia Appendix 2.

Table 1. Population-concept-context (P-C-C) framework.

P-C-C	Definition	Keywords
Population (people with musculoskeletal pain)	Acute (traumatic) or chronic injury related to muscles, bones, joints, tendons, or ligaments problems that cause regional or generalized pain[23] and musculoskeletal disease or conditions as defined by the Global Burden of Disease (rheumatoid arthritis, osteoarthritis, low back pain, neck pain, and gout) [1].	(Musculoskeletal or MSK) injur* or pain* or tear* or ligament* or sprain* or strain* or gout or arthritis or rheumatic arthritis
Concept (digital health)	"The use of information and communications technology in support of health and health-related fields" [10]. Digital health captures eHealth, mobile health, advanced computer sciences, such as big data and AI[a], and the broad scope of telehealth and telemedicine [10].	telemedic* or telehealth or teletriag* or teleconsult* or telecare or tele-care or virtual medicine or virtual care or virtual triage or digital health or digital tool or digital care or digital health technology or AI or artificial intelligence or deep learning or machine learning
Context (triage)	Triage guides the distribution of medical resources to patients when there is a scarcity of health care resources and often refers to a process to allocate, ration, or prioritize patient treatment and is considered first point-of-contact care [4]. Triage can be done by a clinician, patient, or technology (eg, AI) and may involve patients' self-assessment.	self-refer* or self-assess* or self-access* or tele-triage or triage or diagnosis or decision making or symptom checker*

[a]AI: artificial intelligence.

Inclusion Criteria

Studies of adults (aged 18 years and older) with musculoskeletal conditions (≥25% of the sample had to be musculoskeletal-related) that identified and reported a digital health tool designed specifically to triage or diagnose in primary, urgent, or emergency care settings were included. Textbox 1 describes the inclusion and exclusion criteria. We excluded studies that evaluated the effectiveness of virtual assessments. We also excluded studies that used digital health tools for secondary diagnoses (ie, patient had already seen a practitioner and given a diagnosis), as the tools were typically used to manage symptoms and not for primary triage or diagnosis.

Textbox 1. Overview of study selection criteria.

Inclusion criteria
- Adult participants (≥18 years) with a primary complaint of a musculoskeletal condition
- Sample has ≥25% musculoskeletal conditions
- Identifies and reports a digital health tool designed specifically for triage or diagnosis in primary care, urgent care, or emergency settings

Exclusion criteria
- Not English language
- Nonhuman data (eg, vignettes or simulated clinical cases)
- Study design (not original data, eg, review, opinion paper, commentaries, and guidelines)
- Not adult population (all participants aged at least 18 years)
- Not related to a digital health tool (instrument testing or replication or validation studies of clinician assessment to virtual assessment were excluded, no wearable or technology testing was excluded)

Study Selection

Records were collated and uploaded into EndNote (version 20.3; Clarivate Analytics), and duplicates were removed before uploading to Covidence (Veritas Health Innovation) for screening. Pairs of independent reviewers screened all records by title and abstracts, and a third reviewer (CLA) resolved any discrepancies if consensus could not be reached.

At the full-text stage, we first conducted pilot screening, where all reviewers assessed the same 5 full texts. If major discrepancies were identified, we met to review and discuss how to apply the screening criteria in a standardized manner. All full-text papers were reviewed by pairs of independent reviewers. Reasons for exclusion during full-text screening were recorded. Any disagreements between the reviewers at each stage of the selection process were resolved through consensus or by an additional reviewer (CLA) as required.

Data Extraction

Data were extracted from included records independently by pairs of reviewers using a custom data extraction tool designed in Microsoft Excel by the research team. Any disagreements were resolved via consensus. We extracted the following details where available: study characteristics (author, country, sample size, and study aim), participants' demographics (sex, age, and musculoskeletal pain or diagnosis), type of digital tool (name, purpose of tool, and target users), design and development process, platform of tool, tool delivery (eg, clinician or patient self-access), context or care setting in which the tool was used, assessment of performance or accuracy results, and key findings relevant to the review question. Where relevant, authors were contacted once via email to request missing data and to clarify details about the digital health tool.

Data Synthesis

Studies were summarized by study characteristics, digital health tool details, and performance assessment. Descriptive data were summarized as proportions when appropriate. Digital health tools were classified according to the World Health Organization digital health category [10] (eHealth, mHealth, and AI or machine learning), function (ie, triage, diagnosis, or both), care setting in which the tool was used (ie, ED, primary or urgent care, or mixed), research setting (ie, urban, rural, or both), how the tool was administered (eg, self-access or clinician-delivered), technology interface (eg, web-based or app), and intended user (patient-facing, clinician-facing, or both).

Digital tools were rated using the technology readiness level (TRL) and associated technology stage by the first author (LKT) and verified by a second rater [24]. The TRL ranges from 1 to 9, with 1 representing tool conception and 9 representing that the tool is ready to be used in real-world settings [24]. We classified TRL across technology stages (fundamental research, research and development, pilot and demonstration, early adoption, and commercially available) [24].

When available, the performance of the digital health tool was reported by identifying appropriate triage referrals or recommendations or diagnoses compared to a reference standard (eg, physician diagnosis). Measures of performance included diagnostic test accuracy (area under the receiver operator characteristic curve, sensitivity, specificity, positive predictive value, negative predictive value, and likelihood ratio) or reliability measures (internal consistency, test-retest reliability, and intra- or interrater reliability).

Deviations From Protocol

We sought to use the continuous active learning on Covidence to support title and abstract screening [25]. During attempts to calibrate the algorithm, the algorithm did not perform well in identifying relevant papers for this review. We were not confident in screening titles and abstracts with only 1 human reviewer (LKT). Instead, all titles and abstracts and full text records were screened in duplicate by 2 independent human reviewers.

As scoping reviews are an iterative process and aim to assess and evaluate the available evidence, a broad research question often results in a highly sensitive search and less specific records. We made pragmatic decisions and minor amendments to the selection criteria as the review progressed. At the full-text screening stage, studies including a general population had to report ≥25% of the sample being musculoskeletal-related to be included. This cutoff was determined based on studies that indicated the prevalence of musculoskeletal presentations in ED was approximately 25% [26,27].

Results

Overview

The titles and abstracts of 5695 unique records were screened, and 189 papers were reviewed in full (Figure 1). In total, 34 studies met the inclusion criteria (n=37,509 participants across 33 studies, 12,470/37,509, 33% female). The median age was 50 (range 18-91) years. One study did not report the sample size. Sex and age data distribution were missing in 12 and 13 studies, respectively. A list of the studies that were excluded at the full-text stage is presented in Multimedia Appendix 3. Figure 2 charts the data of all 34 studies included by condition, publication year, and sample size of the study, if reported.

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) flowchart.
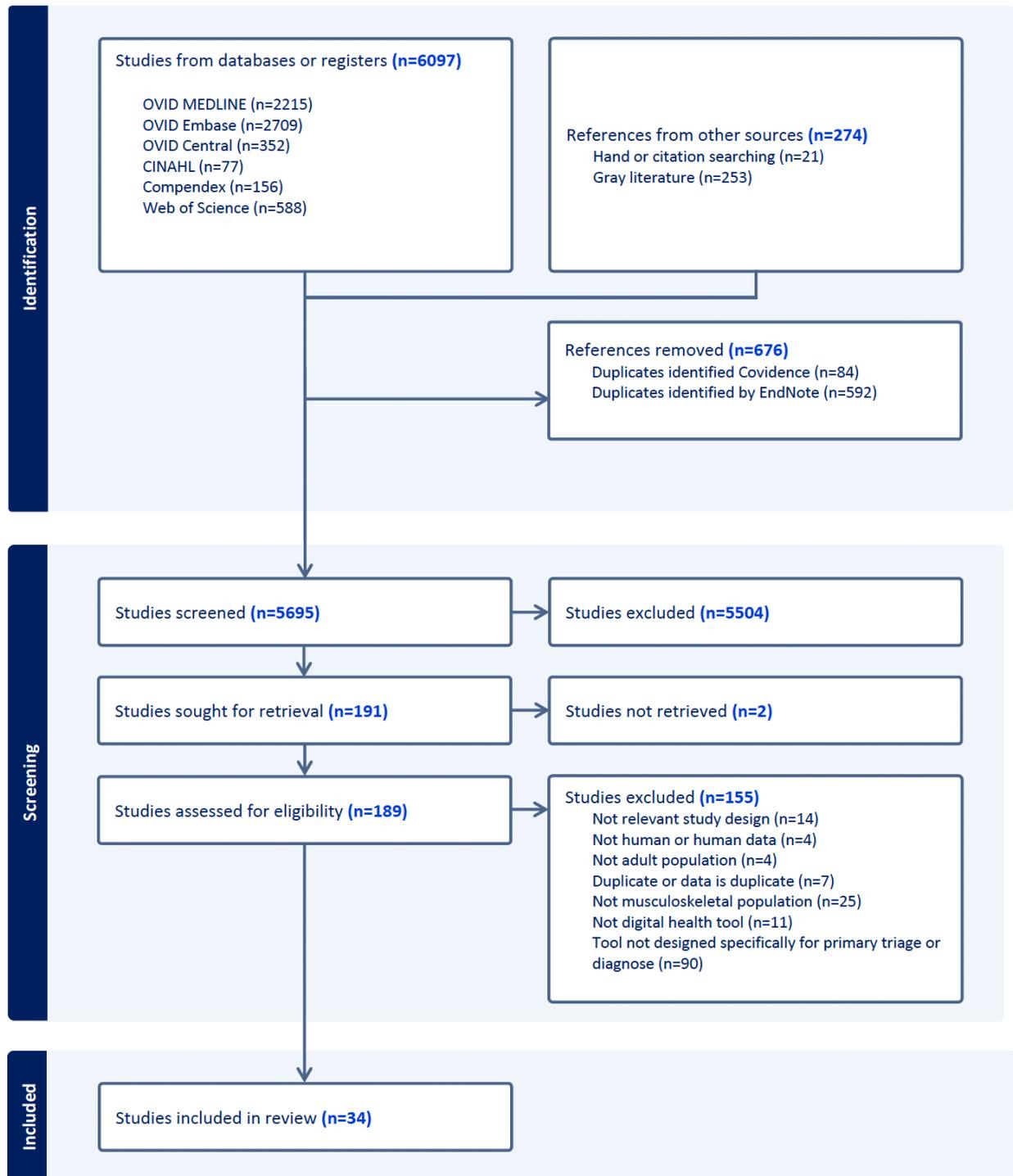
Figure 2. Sources of data charted by publication year and condition.
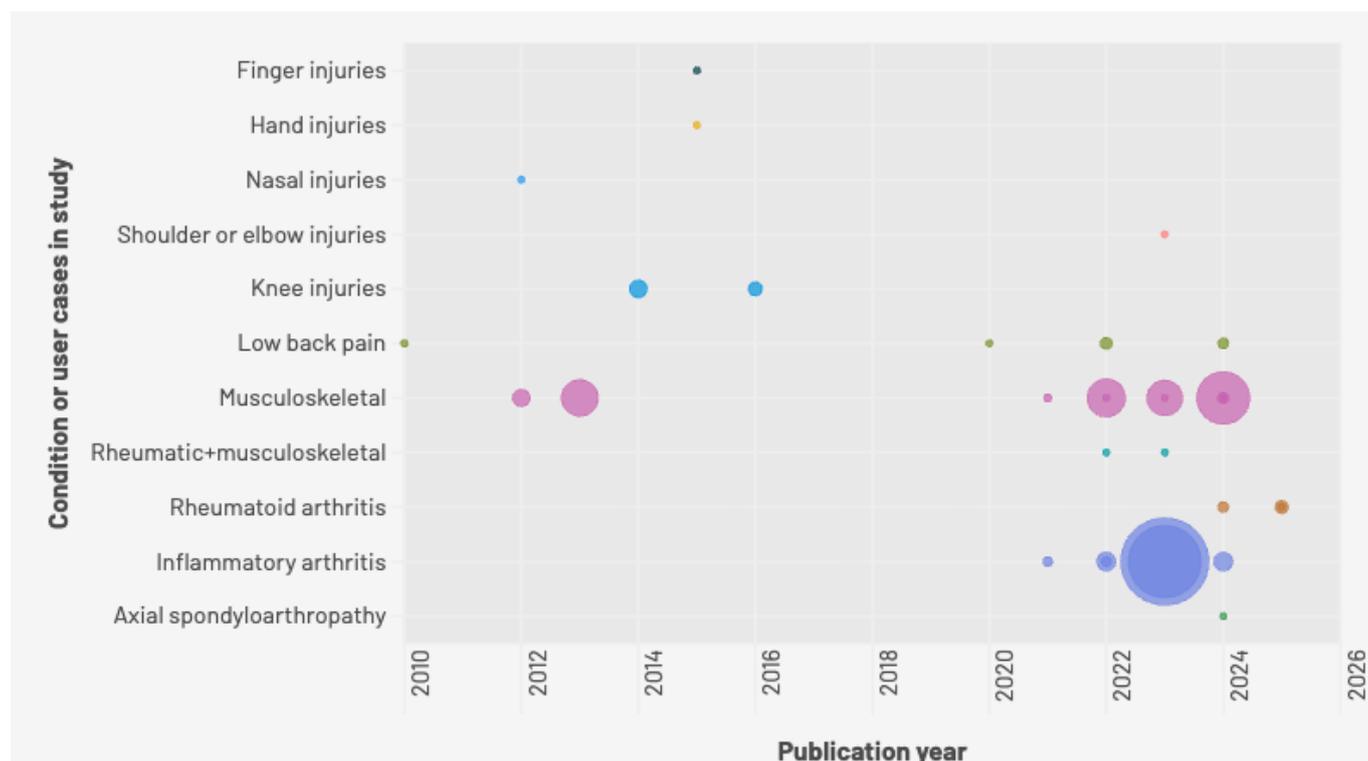
Study Characteristics

In total, 30 of 34 (88%) studies focused on primarily musculoskeletal conditions, and 4 of 34 (12%) studies focused on general populations with a subset of the sample being musculoskeletal conditions. Full study details can be found in Multimedia Appendix 4 [28-61].

In total, 25 of 34 (74%) studies were peer-reviewed, 7 of 34 (21%) records were conference abstracts, and 2 of 34 (6%) were industry case reports. Studies were published between 2010 and 2025. Cross-sectional (10/34, 29%) studies were most common, followed by nonrandomized or quasi-experimental studies (8/34, 24%), randomized controlled trials (5/34, 15%), retrospective cohort (4/34, 12%), mixed or multimethods (3/34), prospective observational cohort (2/34), and case report (2/34, 6%). All studies were conducted in high-income countries (n=1 country not reported), with the majority being from Europe (21/34, 62%) or the United States (7/34, 21%).

Digital health tools were evaluated across various care settings, including ED or urgent care (6/34, 18%), physician-led primary care (6/34, 18%), physiotherapist-led primary care (1/34, 3%), patient self-access (19/34, 56%), and mixed (eg, primary care and ED) settings (2/34, 6%).

Tool Identification and Characteristics

Overview

Inflammatory arthritis (eg, rheumatoid arthritis, gout, and spondyloarthropathy) and arthritis-related conditions were the most common (13/34, 38%) musculoskeletal conditions studied, followed by generic musculoskeletal conditions (11/34, 32%). Others tested digital tools for low back pain (4/34, 12%), knee (2/34, 6%), finger or hand (2/34, 6%), shoulder (1/34, 3%) conditions, and nasal fractures (1/34, 3%).

We identified 16 unique digital health tools (Table 2). In total, 7 studies did not report the name of the digital health tool that was studied or studied a bespoke tool designed for the study where we could not extract the name of the tool. Of the 34 tools, 13 (38%) tools reported that they were designed to "diagnose and triage," 12 (35%) tools were designed to diagnose, and 9 (26%) tools were designed to triage (Table 2). Only 2 tools (Phio [32,39] and Digital Assessment Routing Tool [DART] [51,52]) were designed specifically to triage musculoskeletal conditions. Overall, 3 tools (ADA [37,41,46-49], Buoy Health [33], and WebMD Symptom Checker [40]) were generic health tools that had integrated algorithms to screen musculoskeletal conditions (among other conditions). One tool used OpenAI or LLMs (ChatGPT [29,34,49]) and reported algorithms capable of diagnosing musculoskeletal conditions. Five tools (Rheumatic? [43,45,53,56], Rheport [46-48], RheumConnect [60], ReumAI [38], and Bechterew-check [41]) were condition-specific (ie, designed for rheumatological or inflammatory conditions) with capabilities to differentiate these conditions from other types of musculoskeletal conditions. Two tools were joint-specific (Therapha for low back pain [28] and Virtual Knee Doc for acute knee injuries [30,31]).

Table 2. Characteristics of digital health tools, summarizing purpose, use, technology development, and availability.

Digital health tool name	Purpose of tool[a]	Type of tool[b], intended user, and access level	Tool format	Digital health category	Technology readiness level[c]	Technology readiness assessment[d]	Tool processes[e]	Available to public
ADA [37,41,46-49]	Diagnosis and triage	Symptom checker (patient; self-access)	App	mHealth[f], AI[g] or ML[h]	9	Commercially available	AI	Yes
Bechterew-check [41]	Diagnosis	Symptom checker (patient; self-access)	Web-based	mHealth	7	Pilot and demonstration	Clinical or decision support pathway	Yes (only in German)
Buoy Health [33]	Diagnosis and triage	Symptom checker (patient; self-access)	Web-based or app	eHealth, mHealth, AI or ML	9	Commercially available	AI	Yes
ChatGPT [29,34,49]	Diagnosis	Diagnostic predictor (patient; self-access)	Web-based	eHealth, AI or ML	9	Commercially available	AI	Yes
Digital Assessment Routing Tool [51,52]	Triage	Digital triage[i] (patient; self-access)	App	mHealth	8	Research and development	Clinical or decision support pathway	No
Phio [32,39]	Diagnosis and triage	Symptom checker[j] (patient; self-access)	App	mHealth, AI or ML	9	Commercially available	AI	No, Proprietary
Phone camera [42] (any built-in camera)	Triage	Tele-triage (clinician; clinician-administered)	Phone	mHealth	6	Pilot and demonstration	Clinical or decision support pathway	Yes
PhysioDirect [44]	Diagnosis and triage	Tele-triage (clinician; clinician-administered)	Phone	eHealth	6	Pilot and demonstration	Clinical or decision support pathway	No
Rheumatic? [43,45,53, 56]	Diagnosis and triage	Symptom checker (patient; self-access)	Web-based	eHealth	8	Pilot and demonstration	Clinical or decision support pathway	Yes
Rheport[[37,46-48]	Diagnosis	Symptom checker (patient; self-access)	Web-based	eHealth	9	Early adoption	Clinical or decision support pathway	Yes (only in German)
Therapha [28]	Diagnosis and triage	Digital triage (clinician; clinician-administered)	Web-based	eHealth, AI or ML	9	Commercially available	Clinical or decision support pathway	No (propriety)
TriageXpert Dual Purpose [50]	Triage	Tele-triage (clinician; clinician-administered)	Phone	eHealth	7	Commercially available	Clinical or decision support pathway	No
RheumConnect [60]	Triage	Symptom checker (patient; self-access)	Web-based chatbot	eHealth, AI or ML	6	Pilot and demonstration	AI	No
ReumAI [38]	Triage	Tele-triage (clinician; clinician-administered)	Phone	eHealth, AI or ML	6	Pilot and demonstration	AI	No
Virtual Knee Doc [30, 31]	Diagnosis	Symptom checker (patient; self-access)	Web-based	eHealth	6	Pilot and demonstration	Clinical or decision support pathway	No

Digital health tool name	Purpose of tool[a]	Type of tool[b], intended user, and access level	Tool format	Digital health category	Technology readiness level[c]	Technology readiness assessment[d]	Tool processes[e]	Available to public
WebMD Symptom Checker [40]	Diagnosis	Symptom checker (patient; self-access)	Web-based	eHealth, AI or ML	9	Commercially available	AI	Yes

[a]Triage: provide next steps for care based on symptoms and urgency, may provide preliminary diagnoses but not the objective of the tool. Diagnosis: provide a preliminary diagnosis based on symptoms, which aids to direct next steps in care.

[b]Type of tool: symptom checker: tool designed for patients to enter their symptom data; tele-triage: tool designed to triage using telephone interface; digital triage: tool designed to triage using eHealth or mHealth interface; diagnostic predictor: tool designed to use data to predict diagnosis or triage pathway.

[c]Based on Innovation Canada Technology Readiness Level: rated on a scale of 1-9, where 1=tool conception and 9=tool ready for real-world settings.

[d]Based on Innovation Canada Technology Readiness Stages: fundamental research, research and development, pilot and demonstration, early adoption, commercially available.

[e]Tool processes: AI: tool that uses big data to assign probability to allow for computer-driven decision-making; clinical decision support pathway: predefined decision tree or rule-based algorithms that support clinical decision-making.

[f]mHealth: mobile health.

[g]AI: artificial intelligence.

[h]ML: machine learning.

[i]Self-access within the UK National Health Service.

Intended Users

Most studies (24/34, 71%) reported on digital health tools designed for use by patients, 10 (10/34, 29%) studies targeted tools at clinicians. In total, 19 (19/34, 56%) studies reported on tools that were symptom checkers and were designed to be patient-facing. A total of 10 (29%) studies reported on clinician-facing tools for triage, diagnosis, or diagnostic prediction.

Patient-Facing

We classified ADA, Buoy Health, ChatGPT, Bechterew-check, Rheport, Rheumatic?, RheumConnect, Virtual Knee Doc, DART, and Phio as tools for patients. All used an app or a web-based interface. ADA, Buoy Health, and ChatGPT were tools used for generic health purposes, while the others were designed for specific groups of conditions (ie, rheumatological or musculoskeletal conditions). DART and Phio were designed to integrate with the UK National Health Service. Patients who used DART or Phio had their results forwarded to a primary care team or physiotherapist.

Clinician-Facing

In total, 7 tools were identified for clinicians. Therepha is a clinical decision support system designed for physiotherapists to diagnose and triage low back pain and was piloted in the ED [28]. ReumAI uses tele-triage where a nonphysician staff uses AI-guided telephone interviews to identify diagnoses and potential previsit tests [38]. Triage Xpert Dual Purpose [50] and PhysioDirect [58] were triage tools designed for implementation within specific health systems. One study examined triage of nasal fractures using a built-in camera to triage to the right hospital setting [42], and another used tele-triage to assess whether the ED could be avoided altogether for finger injuries [36]. Most clinician-facing triage tools used tele-triage (ie, phone call) as their interface, except for Therepha, which was a web-based tool. In total, 2 studies leveraged large datasets and AI to predict diagnoses, with 1 study using ChatGPT [29,60].

Performance and Usability

Of the 34 studies identified, 19 (56%) evaluated the performance of the digital health tool (Table 3). The most common definition and method to determine performance was measuring concordance to a clinician diagnosis or recommended triage pathway, often by evaluating sensitivity and specificity. The performance of these tools varied widely and was partly dependent on the context in which they were being used (eg, ED or primary care). Sensitivity ranged from 39% to 91%, and specificity from 23% to 80% [28,30,31, 37,41,45,46,48]. The methods for measuring accuracy were poorly reported, often in the form of proportion of correct triage or diagnoses. Reported accuracy ranged from 33% to 98% across 12 unique tools (n=16 studies) [28,29,34,36-38, 41,45,49,51,59,61]. The accuracy of tools used by patients in tertiary settings (eg, seeking care from a specialist such as an orthopedic surgeon or rheumatologist) was reported as higher than the accuracy of tools used in primary care settings.

For studies that compared digital health tools against each other, ADA was the most common tool used for comparison. When compared to rheumatologists and medical students, ADA was superior to clinician's diagnosis of rheumatic and nonrheumatic conditions, ChatGPT, and Bechterew-check [37,41,49]. ADA was comparable to Rheport for diagnostic accuracy of rheumatic conditions [48]. ChatGPT performed similar to experienced rheumatologists for potential diagnostic accuracy for rheumatic conditions [49].

In total, 4 tools were available in multiple languages (ADA, ChatGPT, Rheumatic?, and WebMD Symptom Checker), and 8 tools were accessible to the public; however, 2 were designed for German speakers (Table 3). Based on the TRL, we classified 6 tools as being at the commercially available stage (ADA, Buoy Health, ChatGPT, Phio, Therapha, and WebMD Symptom Checker).

Figure 3 provides a visualization comparing TRL and performance evaluation for the identified digital health tools (Only 15 of the 16 identified digital tools are reported in this figure. The "phone built in camera" was not graphed.). If the tool did not complete a performance evaluation of the tool, a 0 was given for reported performance (ie, accuracy) on Figure 3. Despite some tools being commercially available, there was a discrepancy in reported performance findings for musculoskeletal conditions.

Table 3. Performance statistics of digital health tools.

Digital health tool and authors	Performance of tool evaluated (yes or no)	Definition used to define tool performance	Condition evaluated	Methods to evaluate performance	Sensitivity (%)	Specificity (%)	Accuracy of tool[a]	Other findings reported (%) (95% CI)
ADA[b]								
Knitza et al (2021) [46]	Yes	Concordance with physician diagnosis	Rheumatic	Sensitivity or specificity, PPV[c], NPV[d]	43	64	NR[e]	PPV 37 (26-48), NPV 69 (60-80)
Knitza et al (2024) [48]	Yes	Concordance with physician diagnosis	Rheumatic	Sensitivity or specificity, PPV, NPV	52	68	NR	PPV or NPV varied depending on whether ADA or Rheport was used first
Graf et al (2022) [37]	Yes	Concordance with identified diagnosis from clinical trial	Rheumatic	Sensitivity or specificity, accuracy	71	64	54% accurately diagnosed same condition	NR
Hannah et al (2024) [41]	Yes	Concordance with discharge summary report	Rheumatic	Sensitivity or specificity, accuracy	39	78	58% accurately diagnosed same condition	NR
Krusche et al (2024) [49]	Yes	Concordance with physician diagnosis	Rheumatic	Proportion	NR	NR	65% accurate for all cases; 71% accurate for cases with IRDs[f]; 61% accurate for non-IRD cases	NR
Bechterew-check								
Hannah et al (2024) [41]	Yes	Concordance with discharge summary report	Axial spondyloarthropathy	Sensitivity or specificity, accuracy	41	53	47% accurately diagnosed same condition	NR
Bespoke tool (no name reported)								
Demmelmaier et al (2010) [35]	No	NT[g]	Low back pain	NT	NT	NT	NT	NR
Martin and Payne (2020) [54]	No	NT	Low back pain	NT	NT	NT	NT	NR
Phillips et al (2012) [55]	No	NT	MSK[h]	NT	NT	NT	NT	NR
Ryan and Grinbergs (2024) [57]	No	NT	MSK	NT	NT	NT	NT	NR
Trivedi et al (2024) [61]	Yes	Concordance with nurse triage	MSK	Proportion	NR	NR	63% accurately triage	NR
Soin et al (2022) [59]	Yes	Concordance with physician diagnosis	Low back pain	NR	NR	NR	72% software predicted correct diagnosis	NR
Buoy Health								
Carmona et al (2022) [33]	No	NT	Generic MSK	NT	NT	NT	NT	NR

Digital health tool and authors	Performance of tool evaluated (yes or no)	Definition used to define tool performance	Condition evaluated	Methods to evaluate performance	Sensitivity (%)	Specificity (%)	Accuracy of tool[a]	Other findings reported (%) (95% CI)
ChatGPT								
Badsha et al (2024) [29]	Yes	Concordance with physician diagnosis	Rheumatic	NR	NR	NR	98% accurate with rheumatologist diagnosis	NR
Daher et al (2023) [34]	Yes	Concordance with physician diagnosis	Shoulder or elbow injuries	NR	NR	NR	93% accurate with surgeon diagnosis; 83% accurate with surgeon management	NR
Krusche et al (2024) [49]	Yes	Concordance with physician diagnosis	Rheumatic	Proportion	NR	NR	35% accurate for all cases; 71% accurate for cases with IRDs; 15% accurate for non-IRD cases	NR
Digital Assessment Routing Tool (DART)								
Lowe et al (2022) [51]	Yes	Concordance with physiotherapist expert	MSK	Proportion	NR	NR	84% DART matched physiotherapist	NR
Lowe et al (2024) [52]	Yes	Concordance with physiotherapist expert	MSK	Intraclass correlation coefficient (ICC)	NR	NR	NR	ICC 0.37 (0.16-0.55)
Phio								
Bond et al (2024) [32]	No	NT	NT	NT	NT	NT	NT	NR
Gymer et al (2023) [39]	No	NT	NT	NT	NT	NT	NT	NR
Phone Camera								
Hara et al (2015) [42]	Yes	Accuracy of triage recommendations	Finger injuries	NR	NR	NR	NR	NR
PhysioDirect								
Kelly et al (2021) [44]	No	NT	NT	NT	NT	NT	NT	NR
Rheport								
Knitza et al (2021) [46]	Yes	Concordance with physician diagnosis	Rheumatic	Sensitivity or specificity, PPV, NPV	54	52	NR	PPV 35 (25-47); NPV 70 (58-79)
Knitza et al (2024) [48]	Yes	Concordance with physician diagnosis	Rheumatic	Sensitivity or specificity, PPV, NPV	62	47	NR	PPV or NPV varied depending on whether ADA or Rheport was used first
Rheumatic?[i]								

Digital health tool and authors	Performance of tool evaluated (yes or no)	Definition used to define tool performance	Condition evaluated	Methods to evaluate performance	Sensitivity (%)	Specificity (%)	Accuracy of tool[a]	Other findings reported (%) (95% CI)
Knevel et al (2022) [45]	Yes	Concordance with physician diagnosis/treatment recommendation	Rheumatic	Sensitivity or specificity, AUC-ROC[j]	67	72	AUC-ROC 75 (95% CI 62-89)	NR
Qin et al (2024) [56]	No	NT	Rheumatic	NT	NT	NT	NT	NR
Lundberg et al (2023) [53]	No	NT	Rheumatic	NT	NT	NT	NT	NR
Jakobi et al (2025) [43]	No	NT	Rheumatic	NT	NT	NT	NT	NR
ReumAI								
Gómez-Centeno et al (2025) [38]	Yes	Concordance with physician diagnosis	Rheumatic	NR	NR	NR	53% accurate with rheumatologists	NR
RheumConnect								
Tan et al (2023) [60]	No	NT	Rheumatic	NT	NT	NT	NT	NT
Therapha								
Badahman et al (2024) [28]	Yes	Concordance with MRI findings	Low back pain	Sensitivity or specificity, PPV, NPV, ROC[k]	88	80	ROC 0.84 (95% CI 0.6-1.0; $P=.001$)	PPV 99 (25-47); NPV 27 (58-79)
Triage Xpert Dual Purpose								
Li et al (2023) [50]	No	NT	MSK	NT	NT	NT	NT	NT
Virtual Knee Doc								
Bisson et al (2014) [30]	Yes	Concordance with physician diagnosis	Knee injuries	Sensitivity or specificity	89	27	NR	NT
Bisson et al (2016) [31]	Yes	Concordance with physician diagnosis	Knee injuries	Sensitivity or specificity	91	23	NR	NT
WebMD Symptom Checker								
Hageman et al (2015) [40]	Yes	Concordance with physician diagnosis	Hand injuries	Proportion	NR	NR	33% accurate with hand surgeon diagnosis	NT

[a]Reported values from the study and not interpretation of authors.
[b]Findings reported from ADA diagnosis 1 (D1) in study.
[c]PPV: positive predictive value.
[d]NPV: negative predictive value.
[e]NR: not reported.
[f]IRD: inflammatory rheumatic disease.
[g]NT: not tested.
[h]MSK: musculoskeletal.
[i]Findings reported from dataset A in study.
[j]AUC-ROC: area under the receiver operating curve.
[k]ROC: receiver operating curve.

Figure 3. Visualization comparing TRL and the highest reported performance evaluation across identified digital health tools. NR: not reported; TRL: technology readiness level.
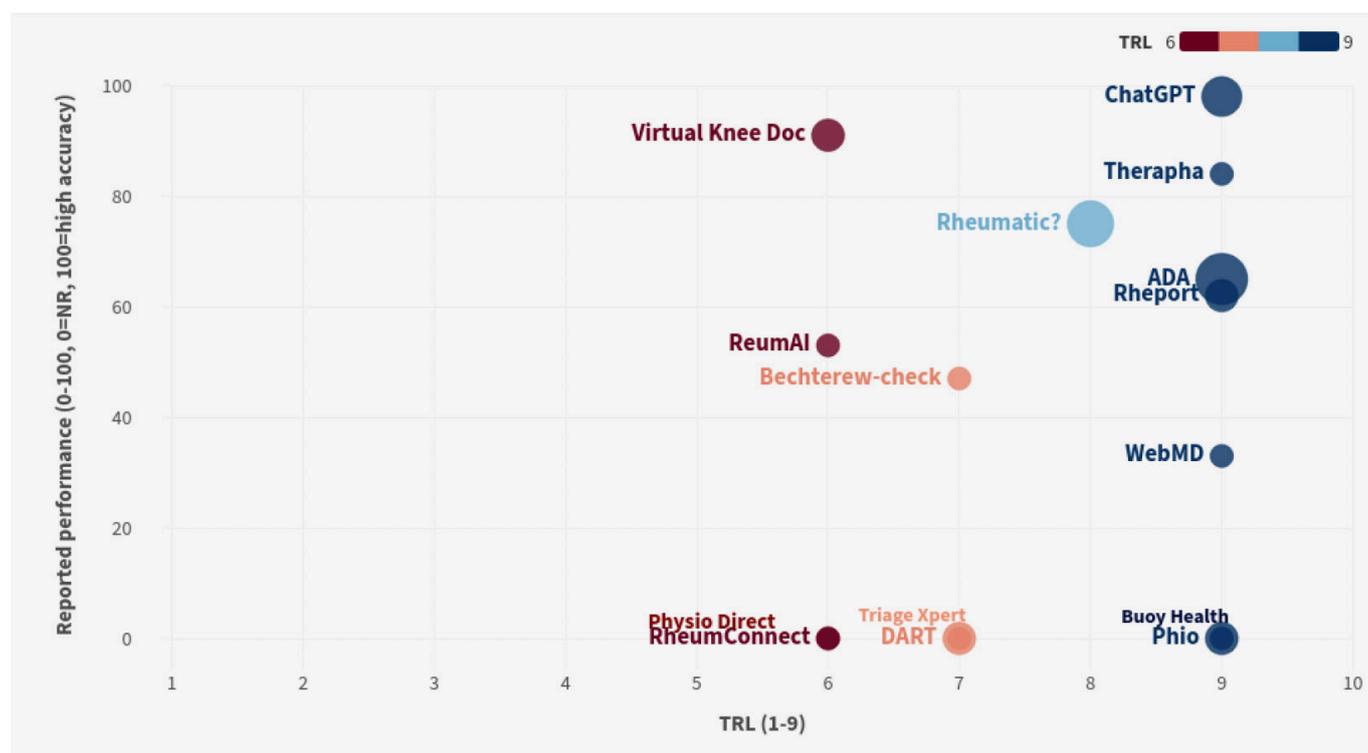
Discussion

Principal Findings

We aimed to identify and describe the available tools for triaging and diagnosing musculoskeletal conditions in primary, urgent, and emergency settings. Based on a synthesis of 34 studies and data from 16 different digital health tools, there were no digital health tools with sufficient evidence to support effective triage and diagnosis of musculoskeletal conditions across these settings. Approximately half of these tools were available to the public. Not all tools were available in English, with 2 tools only available in German (Bechterew-check and Rheport). The most frequently studied digital health tool was ADA (n=5), followed by Rheumatic? (n=4), then ChatGPT (n=3). Only 2 tools (DART and Phio) were purposely developed for screening musculoskeletal conditions. Both tools are not currently available outside of the UK's National Health Service. We were surprised to find so few digital health tools targeting musculoskeletal conditions, given the substantial global burden of musculoskeletal conditions [1]. Notably, rheumatological or inflammatory arthritis was the most prevalent musculoskeletal condition studied, despite low back pain being the most common musculoskeletal condition seen in ED and primary care settings [62]. We identified 4 studies that included digital health tools targeting low back pain, but only 1 of these reported which tool was used (Therapha). Our findings reflect the discordance of research across digital health technology and the current health landscape. Many tools were inaccessible or not designed for practical use in managing musculoskeletal pain, the most burdensome conditions seen in primary care.

Our secondary objective was to summarize the performance and accuracy of the included digital health tools. Approximately 50% of the studies evaluated the performance of a digital health tool. Apart from ChatGPT, most generic digital health tools (eg, ADA and WebMD Symptom Checker) reported poor accuracy (often less than 50% accuracy in identifying the correct diagnosis compared to clinicians) for musculoskeletal conditions [37,40,48]. Despite the use of ChatGPT by the public as a symptom checker [15], ChatGPT's accuracy for diagnosing musculoskeletal and rheumatic conditions was variable, ranging from 33% to 98% [29,34,49]. We suggest that further research is needed before considering ChatGPT as an accurate diagnostic or screening tool. Tools that were designed to diagnose peripheral or spinal musculoskeletal conditions (eg, low back pain or knee injuries) appear to be more promising with high sensitivity (88%-91%) [28,31]. Finally, tools designed specifically to triage (rather than diagnose) musculoskeletal conditions (ie, Phio and DART) demonstrated the best performance. Recent findings published on DART and Phio indicate that these tools have high agreement (>90%) with expert physiotherapist recommendations on next care pathways [63, 64]. However, the heterogeneity across evaluation methods highlights the importance of standardized development and evaluation frameworks to ensure that digital triage tools for musculoskeletal conditions are accurate, transparent, and safe before integrated into clinical settings and workflows.

Not Yet Ready for Prime Time

One of the key findings of our review is that some tools are commercially available and integrated into health systems for musculoskeletal screening without robust methodological

evaluation or reporting. Premature implementation raises concerns, particularly given the risk of misdirecting patients or delaying appropriate care. Before being adopted at scale, digital triage tools must demonstrate value in real-world settings and meet minimum standards for safety, accuracy, and usability. However, many studies evaluating these tools lack transparent reporting, making it difficult to assess how performance claims were derived. Studies reporting high accuracy of their digital tools often had poor transparency or a lack of details on their tool evaluation. We suggest caution with interpreting these digital health tools as ready for public use without further evaluation. It is also unclear how tools that use LLMs operate. These networks are often termed "black boxes" due to the inability to explain how these systems achieve their output [65].

Our findings have been confirmed by a recent study that evaluated diagnostic accuracy and clinical reasoning using 6 different generative AI (LLMs) for rheumatic diagnoses [66]. Despite the LLMs reporting high diagnostic accuracy (~80%), all models reported subpar clinical reasoning quality (eg, explaining reasons for supporting diagnoses) [66]. These findings underscore the importance of digital health tools requiring both high diagnostic accuracy alongside transparent algorithms to help to explain the logic behind the tool's decision. To improve transparency and enable reproducibility, it is important to establish standards for incorporating ethical AI in digital health. Without transparency in how tools were developed, or in the algorithms used, it is unclear whether the tools are safe for the public to use.

The only digital health tool with robust evaluation of its performance was the generic health app, ADA, which is a Conformité Européenne–certified medical product [46,48]. ADA's performance was inconsistent across the studies, and ADA correctly identified the musculoskeletal condition or triage option in fewer than half of the cases [37,41,46,48, 49]. Condition-specific digital health tools (Rheport [46,48], Rheumatic? [45], Virtual Knee Doc [30,31], Therapha [28], and ReumAI [38]) performed slightly better. The reported accuracy was higher in these tools, especially if these tools were implemented in tertiary care settings (outside of the ED or primary care). We are not aware of an acceptable threshold for performance (ie, accuracy) for digital health tools. However, we recommend implementing tools that are at least more accurate than flipping a coin and provide consistent results across different study contexts or musculoskeletal conditions.

AI-driven tools, like ADA or ChatGPT, may perform better than clinician decision support systems or physicians or rheumatologists in diagnosing rheumatic conditions [49, 67]. Integrating digital health tools in tandem with other nonspecialist professions (eg, general practitioners and allied health professionals) could help guide patients to their next care steps as they wait for specialists (eg, rheumatologists) or avoid unnecessary visits to specialists or other care providers. AI-driven tools that have included diagnostic findings (eg, imaging, clinical symptoms or signs, and bloodwork) have superior diagnostic accuracy to other AI models [41,67,68]. Until robust stand-alone digital health tools are developed

(ie, a symptom checker that can be used independently by patients), combining digital health tools and clinician feedback may be the best method to streamline diagnosis and care in complex cases while providing timely care for common musculoskeletal conditions.

Several frameworks for evaluating digital health tools have been proposed [69]. A recent scoping review identified 12 key domains—ranging from tool description and content to safety, clinical effectiveness, and efficacy—across 95 frameworks that developers and researchers can draw on [69]. However, the heterogeneity reflects a broader challenge: many digital health tools span multiple categories (eg, eHealth or mHealth tools incorporating AI), making classification inconsistent and evaluation difficult. Advancing this field requires standardized terminology, harmonized testing and evaluation frameworks, and clear reporting guidelines—crucial steps to ensure both progress and patient safety.

Why Would a Digital Health Tool Do a Poor Job at Screening Musculoskeletal Conditions?

Through the process of screening studies for inclusion into our review, we found definitions of musculoskeletal conditions that were vague and varied widely. Definitions of "musculoskeletal" are often limited to orthopedic conditions or pain related to musculoskeletal structures [1,23]. However, musculoskeletal conditions are a complex category involving heterogeneous conditions, such as rheumatological or inflammatory arthritis or gout, that are not typically grouped as musculoskeletal in clinical practice. We relied on a broad definition to capture specific musculoskeletal conditions (eg, rheumatological conditions, arthritis, and gout) and pain related to musculoskeletal structures (eg, sprains and strains).

There is nuance in how triage would be conducted for acute versus chronic musculoskeletal conditions, including screening questions related to condition pathophysiology, subjective history, pattern of symptoms, and disability (eg, red flags), which might explain some of the variability in performance metrics of different digital tools [70]. Early diagnosis and treatment planning is often iterative for those with musculoskeletal conditions and varies depending on the condition. For example, targeted medication plays a vital role in managing rheumatological conditions [71], whereas some orthopedic conditions are managed with exercise and minimal pharmacological interventions [2]. This complexity will impact triage algorithms by influencing treatment recommendations (eg, who the patient should see) and timing of care (eg, urgent or wait-and-see). Therefore, tools that have high accuracy (ie, good performance) for triaging and diagnosing general health conditions may not necessarily have the same effectiveness when applied to musculoskeletal conditions.

Digital health tools may perform poorly at screening because of user error relating to symptom data entry and patient interaction with the tool. One solution to this is adding

more key information (eg, diagnostic tests) to an AI-driven model to improve the diagnostic accuracy of the model [67]. We also suggest future work to involve patient end users to develop and refine digital health tools. Most digital health tool algorithms are derived from clinicians' clinical reasoning, which may not follow the same thought process as a patient. In a recent qualitative study exploring how patients should be engaged in AI application to health care, patients felt that the priorities of researchers, particularly for AI tools, were to improve efficiency and effectiveness of care [72]. In contrast, patients were more interested in using AI to address issues related to accessing health care [72]. Patients should be involved early in the design and development phases to enhance the usability and understandability of digital health tools. However, patient perspectives are often included only after the digital health tool is designed. We argue that engaging patients early in the development process, such as developing the AI algorithms, may yield more acceptable and usable digital health tools.

It is unlikely that a "one size fits all" digital health tool can effectively diagnose and triage all musculoskeletal conditions. Most patient-facing tools in our review were web- or app-based tools in the form of generic symptom checkers. ChatGPT has an accessible interface and is relatively easy to use [15]. Clinician-facing tools may benefit from greater complexity or condition specificity, depending on the context in which the tools will be implemented. Instead of an either-or—general or condition-specific—we advocate for designers to consider their design goal (ie, triage or diagnosis) and intended user (ie, patient or clinician), which may improve accuracy in digital health tools for musculoskeletal conditions.

Move (Relatively) Fast, and Try Not to Break Things

The field of digital health is growing and changing rapidly. Many health systems have been forced to move toward implementing digital health, particularly AI-driven tools, without being afforded adequate time and resources to consider safety, effectiveness, or downstream consequences [13]. This may be in part due to social and political imperatives to set key performance (productivity) indicators, transition of health care services, and drive toward greater and faster innovation. We suggest that such a climate could be dangerous for health care, especially if digital health implementation continues without adequate evidence, as our findings highlight.

There is a place for digital health triage tools used by patients and clinicians in the current health care context. Self-referral and symptom checkers can be effective for musculoskeletal conditions and to support patients' access to care, particularly when patients do not have a consistent primary care team or provider [11]. Acute care clinics using a self-referral form found that patients with musculoskeletal conditions were accurate at self-referring, used less health care, and incurred fewer costs [73]. Emerging evidence also indicates that patients are using LLMs such as ChatGPT to make health care decisions, and it appears that the general public is accepting of using AI for health care advice and psychological support [15]. However, more research is needed to ensure that patients presenting with musculoskeletal conditions have a safe, accurate, and well-designed tool to direct them to the best care for their situation. Digital health tools also need to be designed to suit diverse populations, including those with low health literacy and limited digital literacy.

Future Considerations and Clinical Implications

While there is a breadth of studies available for digital health and digital triage, we identified the following knowledge gaps: (1) reporting and transparency on digital health tool development must improve, (2) evaluating digital health tools needs a standard approach, (3) studying the accuracy of triage recommendations requires robust prospective studies, and (4) implementing musculoskeletal-focused digital health tools for first point-of-contact care requires attention.

Despite the absence of digital health tools for triage of musculoskeletal conditions, we are aware of other tools in development, such as SupportPrim [74], which might fill some of the knowledge gaps for health care providers. Our findings do not provide conclusive evidence to support using digital health tools to accurately screen musculoskeletal conditions in many health settings. We recommend that clinicians use these digital health tools as an adjunct to help guide patients, particularly when used as a symptom checker, but to still defer to sound clinical judgment and help patients understand the limitations of the tools.

Limitations

Although we used a thorough search of published and unpublished data, it is possible that we have missed relevant digital health tools or papers. We set a sample threshold of at least 25% of the sample population with musculoskeletal conditions, and this may have resulted in us missing some studies (eg, studies that were just below the threshold were excluded). The threshold was intended to maximize external validity [26,27]. Our goal was to identify tools that were primarily designed to triage or diagnose (vs manage) musculoskeletal conditions. Therefore, we excluded studies and tools that were designed for self-management, even if they included a symptom checker. This led us to exclude studies that used tools for secondary triage or diagnosis (ie, used by patients who had a diagnosis or had already been seen in a primary or emergency setting) as we wanted to capture tools that could be used at the first point-of-contact. We identified some potential musculoskeletal-specific digital health tools that could be used for secondary triage or diagnosis (Multimedia Appendix 5). While we attempted to report on the performance and accuracy of the tools identified, some tools pooled data from the entire population (ie, not musculoskeletal only). Therefore, the findings may under- or overestimate the accuracy of the tool for musculoskeletal conditions. This again points to the need to design musculoskeletal-specific tools and carefully evaluate their performance.

Conclusions

The rapid growth of AI and digital health solutions is transforming health care systems worldwide, with increasing interest in automating triage and diagnosis. However, our review shows that musculoskeletal conditions remain a blind spot: few tools were specifically designed for this purpose, and most performed poorly when applied to musculoskeletal populations. Despite commercial availability and implementation in some settings, the evidence base was weak, and tool performance was inconsistent and opaque. Health systems and clinicians should exercise caution before integrating these tools into care pathways. Musculoskeletal-specific digital tools developed through transparent, standardized processes are urgently needed to ensure safety, clinical value, and trustworthiness.

Data Availability

The datasets generated or analyzed during this study are available in the Open Science Framework repository [22].

Authors' Contributions

Conceptualization: LKT, CLA, JGW (equal)
Data curation (database searching): DG (lead), LKT, CLA (supporting)
Investigation: LKT, JGW, RV, EL, JLC, EW, CS, CLA
Methodology: LKT, CLA (equal)
Formal analysis: LKT (lead), CLA, JGW, RV (supporting)
Project administration: LKT (lead), CLA (supporting)
Visualization: LKT (lead), JGW, CLA (supporting)
Writing—original draft: LKT (lead), CLA (supporting)
Writing—review and editing: All authors

Conflicts of Interest

None declared.

Multimedia Appendix 1

OVID MEDLINE full search strategy.
[DOCX File (Microsoft Word File), 50 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Gray literature search strategy and results.
[DOCX File (Microsoft Word File), 27 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Studies excluded at full-text stage.
[DOCX File (Microsoft Word File), 74 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Characteristics of included studies, summarizing design, demographics, and digital tool features.
[DOCX File (Microsoft Word File), 57 KB-Multimedia Appendix 4]

Multimedia Appendix 5

Digital health tools for secondary triage or diagnosis.
[DOCX File (Microsoft Word File), 25 KB-Multimedia Appendix 5]

Checklist 1

PRISMA-ScR checklist.
[PDF File (Adobe File), 249 KB-Checklist 1]

References

1. Gill TK, Mittinty MM, March LM, et al. Global, regional, and national burden of other musculoskeletal disorders, 1990–2020, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. Lancet Rheumatol. Nov 2023;5(11):e670-e682. [doi: 10.1016/S2665-9913(23)00232-1]

2. Lin I, Wiles L, Waller R, et al. What does best practice care for musculoskeletal pain look like? Eleven consistent recommendations from high-quality clinical practice guidelines: systematic review. Br J Sports Med. Jan 2020;54(2):79-86. [doi: 10.1136/bjsports-2018-099878] [Medline: 30826805]

3. Wrightson J, Truong LK, Haagaard A, Ardern CL. Estimating the prevalence of low acuity musculoskeletal pain in the emergency department. Presented at: Canadian for Health Services and Policy Research (CHSPR) 2025 Conference Abstract; Mar 3-4, 2025; Vancouver, Canada.

4. Iserson KV, Moskop JC. Triage in medicine, part I: concept, history, and types. Ann Emerg Med. Mar 2007;49(3):275-281. [doi: 10.1016/j.annemergmed.2006.05.019] [Medline: 17141139]

5. Joseph C, Morrissey D, Abdur-Rahman M, Hussenbux A, Barton C. Musculoskeletal triage: a mixed methods study, integrating systematic review with expert and patient perspectives. Physiotherapy. Dec 2014;100(4):277-289. [doi: 10.1016/j.physio.2014.03.007] [Medline: 25242531]

6. Erku D, Khatri R, Endalamaw A, et al. Digital health interventions to improve access to and quality of primary health care services: a scoping review. Int J Environ Res Public Health. Sep 28, 2023;20(19):19. [doi: 10.3390/ijerph20196854] [Medline: 37835125]

7. Ibrahim MS, Mohamed Yusoff H, Abu Bakar YI, Thwe Aung MM, Abas MI, Ramli RA. Digital health for quality healthcare: a systematic mapping of review studies. Digit Health. 2022;8:20552076221085810. [doi: 10.1177/20552076221085810] [Medline: 35340904]

8. Chambers D, Cantrell AJ, Johnson M, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. BMJ Open. Aug 1, 2019;9(8):e027743. [doi: 10.1136/bmjopen-2018-027743] [Medline: 31375610]

9. Global strategy on digital health 2020-2025. World Health Organization. 2021. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [Accessed 2025-10-24]

10. Recommendations on digital interventions for health system strengthening—executive summary. World Health Organization. 2019. URL: https://www.who.int/publications/i/item/WHO-RHR-19.8 [Accessed 2025-01-05]

11. Babatunde OO, Bishop A, Cottrell E, et al. A systematic review and evidence synthesis of non-medical triage, self-referral and direct access services for patients with musculoskeletal pain. PLOS ONE. 2020;15(7):e0235364. [doi: 10.1371/journal.pone.0235364] [Medline: 32628696]

12. Pairon A, Philips H, Verhoeven V. A scoping review on the use and usefulness of online symptom checkers and triage systems: how to proceed? Front Med (Lausanne). 2022;9:1040926. [doi: 10.3389/fmed.2022.1040926] [Medline: 36687416]

13. Tyler S, Olis M, Aust N, et al. Use of artificial intelligence in triage in hospital emergency departments: a scoping review. Cureus. May 2024;16(5):e59906. [doi: 10.7759/cureus.59906] [Medline: 38854295]

14. Wallace W, Chan C, Chidambaram S, et al. The diagnostic and triage accuracy of digital and online symptom checker tools: a systematic review. NPJ Digit Med. Aug 17, 2022;5(1):118. [doi: 10.1038/s41746-022-00667-w] [Medline: 35977992]

15. Shahsavar Y, Choudhury A. User intentions to use ChatGPT for self-diagnosis and health-related purposes: cross-sectional survey study. JMIR Hum Factors. May 17, 2023;10:e47564. [doi: 10.2196/47564] [Medline: 37195756]

16. Peters MDJ, Godfrey C, McInerney P, et al. Best practice guidance and reporting items for the development of scoping review protocols. JBI Evid Synth. Apr 1, 2022;20(4):953-968. [doi: 10.11124/JBIES-21-00242] [Medline: 35102103]

17. Pollock D, Peters MDJ, Khalil H, et al. Recommendations for the extraction, analysis, and presentation of results in scoping reviews. JBI Evid Synth. Mar 1, 2023;21(3):520-532. [doi: 10.11124/JBIES-22-00123] [Medline: 36081365]

18. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med. Oct 2, 2018;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]

19. Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. Syst Rev. Jan 26, 2021;10(1):39. [doi: 10.1186/s13643-020-01542-z] [Medline: 33499930]

20. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol. Feb 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]

21. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci. Sep 20, 2010;5:69. [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]

22. Truong LK, Lui E, Wrightson J, et al. Digital health tools used to triage musculoskeletal pain in primary, urgent and emergency settings: a scoping review. Open Science Framework. 2024. URL: https://osf.io/y5rp7/overview [Accessed 2026-01-05]

23. Smith E, Hoy DG, Cross M, et al. The global burden of other musculoskeletal disorders: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis. Aug 2014;73(8):1462-1469. [doi: 10.1136/annrheumdis-2013-204680] [Medline: 24590181]

24. Technology Readiness Level (TRL) Assessment Tool. Government of Canada. 2021. URL: https://ised-isde.canada.ca/site/clean-growth-hub/en/technology-readiness-level-trl-assessment-tool [Accessed 2025-05-24]

25. Walton A. December 2022—title and abstract screening using machine learning. Covidence. 2022. URL: https://www.covidence.org/blog/release-notes-december-2022-machine-learning [Accessed 2024-07-11]

26. Bird S, Thompson C, Williams KE. Primary contact physiotherapy services reduce waiting and treatment times for patients presenting with musculoskeletal conditions in Australian emergency departments: an observational study. J Physiother. Oct 2016;62(4):209-214. [doi: 10.1016/j.jphys.2016.08.005] [Medline: 27637771]

27. Gagnon R, Perreault K, Berthelot S, et al. Direct-access physiotherapy to help manage patients with musculoskeletal disorders in an emergency department: results of a randomized controlled trial. Acad Emerg Med. Aug 2021;28(8):848-858. [doi: 10.1111/acem.14237] [Medline: 33617696]

28. Badahman F, Alsobhi M, Alzahrani A, et al. Validating the accuracy of a patient-facing clinical decision support system in predicting lumbar disc herniation: diagnostic accuracy study. Diagnostics (Basel). Aug 26, 2024;14(17):1870. [doi: 10.3390/diagnostics14171870] [Medline: 39272655]

29. Badsha HM, Khan B, Harifi G, J A, Raman S. AB1488 is the future of rheumatology here? A study of a proprietary rule engine and artificial intelligence GPT4 (AI GPT4) for initial evaluation of rheumatology cases. Ann Rheum Dis. Jun 2024;83(Suppl 1):2112. [doi: 10.1136/annrheumdis-2024-eular.1942] [Medline: 644868572]

30. Bisson LJ, Komm JT, Bernas GA, et al. Accuracy of a computer-based diagnostic program for ambulatory patients with knee pain. Am J Sports Med. Oct 2014;42(10):2371-2376. [doi: 10.1177/0363546514541654] [Medline: 25073597]

31. Bisson LJ, Komm JT, Bernas GA, et al. How accurate are patients at diagnosing the cause of their knee pain with the help of a web-based symptom checker? Orthop J Sports Med. Feb 2016;4(2):2325967116630286. [doi: 10.1177/2325967116630286] [Medline: 26962542]

32. Bond C, Guard M, Grinbergs P. Case report: digital musculoskeletal triage and rehabilitation tools enhance accessibility, user experience and outcomes in mechanical knee pain. Physiotherapy. Jun 2024;123(Suppl 1):e115. [doi: 10.1016/j.physio.2024.04.143]

33. Carmona KA, Chittamuru D, Kravitz RL, Ramondt S, Ramírez AS. Health information seeking from an intelligent web-based symptom checker: cross-sectional questionnaire study. J Med Internet Res. Aug 19, 2022;24(8):e36322. [doi: 10.2196/36322] [Medline: 35984690]

34. Daher M, Koa J, Boufadel P, Singh J, Fares MY, Abboud JA. Breaking barriers: can ChatGPT compete with a shoulder and elbow specialist in diagnosis and management? JSES Int. Nov 2023;7(6):2534-2541. [doi: 10.1016/j.jseint.2023.07.018] [Medline: 37969495]

35. Demmelmaier I, Denison E, Lindberg P, Asenlöf P. Physiotherapists' telephone consultations regarding back pain: a method to analyze screening of risk factors. Physiother Theory Pract. Oct 2010;26(7):468-475. [doi: 10.3109/09593980903433938] [Medline: 20649497]

36. Dias L, Maughan E, Kisha A, Moorthy R. Telephone triage in the management of patients with nasal injuries [Abstract]. Clin Otolaryngol. 2012;37:45. [doi: 10.1111/j.1749-4486.2012.02517.x] [Medline: 71023181]

37. Gräf M, Knitza J, Leipe J, et al. Comparison of physician and artificial intelligence-based symptom checker diagnostic accuracy. Rheumatol Int. Dec 2022;42(12):2167-2176. [doi: 10.1007/s00296-022-05202-4] [Medline: 36087130]

38. Gómez-Centeno A, Sabaris-Vilas M, Garcia-Sancho F, Segura-Sanchez J. POS0883 Optimizing rheumatology consultations with artificial intelligence: insights from the ReumAI pilot study. Ann Rheum Dis. Jun 2025;84(Suppl 1):1018. [doi: 10.1016/j.ard.2025.06.238]

39. Gymer M, Guard M, Grinbergs P. A case report: digital musculoskeletal triage and rehabilitation tools improve outcomes and offer a positive experience for lower back pain. JMIR Bioinform Biotechnol. Preprint posted online on Oct 20, 2022. [doi: 10.2196/preprints.43686]

40. Hageman M, Anderson J, Blok R, Bossen JKJ, Ring D. Internet self-diagnosis in hand surgery. HAND (N Y). Sep 2015;10(3):565-569. [doi: 10.1007/s11552-014-9707-x] [Medline: 26330798]

41. Hannah L, von Sophie R, Gabriella RM, et al. Stepwise asynchronous telehealth assessment of patients with suspected axial spondyloarthritis: results from a pilot study. Rheumatol Int. Jan 2024;44(1):173-180. [doi: 10.1007/s00296-023-05360-z] [Medline: 37316631]

42. Hara T, Nishizuka T, Yamamoto M, Iwatsuki K, Natsume T, Hirata H. Teletriage for patients with traumatic finger injury directing emergency medical transportation services to appropriate hospitals: a pilot project in Nagoya City, Japan. Injury. Jul 2015;46(7):1349-1353. [doi: 10.1016/j.injury.2015.02.022] [Medline: 25799472]

43. Jakobi S, Boy K, Wagner M, et al. Rheumatic? A diagnostic decision support tool for individuals suspecting rheumatic diseases: mixed-methods usability and acceptability study. BMC Rheumatol. May 23, 2025;9(1):59. [doi: 10.1186/s41927-025-00507-w] [Medline: 40410901]

44. Kelly M, Higgins A, Murphy A, McCreesh K. A telephone assessment and advice service within an ED physiotherapy clinic: a single-site quality improvement cohort study. Arch Physiother. Feb 8, 2021;11(1):4. [doi: 10.1186/s40945-020-00098-4] [Medline: 33550990]

45. Knevel R, Knitza J, Hensvold A, et al. Rheumatic?—A digital diagnostic decision support tool for individuals suspecting rheumatic diseases: a multicenter pilot validation study. Front Med. 2022;9. [doi: 10.3389/fmed.2022.774945] [Medline: 2016520483]

46. Knitza J, Mohn J, Bergmann C, et al. Accuracy, patient-perceived usability, and acceptance of two symptom checkers (Ada and Rheport) in rheumatology: interim results from a randomized controlled crossover trial. Arthritis Res Ther. Apr 13, 2021;23(1):112. [doi: 10.1186/s13075-021-02498-8] [Medline: 33849654]

47. Knitza J, Muehlensiepen F, Ignatyev Y, et al. Patient's perception of digital symptom assessment technologies in rheumatology: results from a multicentre study. Front Public Health. 2022;10. [doi: 10.3389/fpubh.2022.844669] [Medline: 637486321]

48. Knitza J, Tascilar K, Fuchs F, et al. Diagnostic accuracy of a mobile AI-based symptom checker and a web-based self-referral tool in rheumatology: multicenter randomized controlled trial. J Med Internet Res. Jul 23, 2024;26:e55542. [doi: 10.2196/55542] [Medline: 39042425]

49. Krusche M, Callhoff J, Knitza J, Ruffer N. Diagnostic accuracy of a large language model in rheumatology: comparison of physician and ChatGPT-4. Rheumatol Int. Feb 2024;44(2):303-306. [doi: 10.1007/s00296-023-05464-6] [Medline: 37742280]

50. Li KY, Kim PS, Thariath J, Wong ES, Barkham J, Kocher KE. Standard nurse phone triage versus tele-emergency care pilot on Veteran use of in-person acute care: an instrumental variable analysis. Acad Emerg Med. Apr 2023;30(4):310-320. [doi: 10.1111/acem.14681] [Medline: 36757685]

51. Lowe C, Browne M, Marsh W, Morrissey D. Usability testing of a digital assessment routing tool for musculoskeletal disorders: iterative, convergent mixed methods study. J Med Internet Res. Aug 30, 2022;24(8):e38352. [doi: 10.2196/38352] [Medline: 36040787]

52. Lowe C, Sephton R, Marsh W, Morrissey D. Evaluation of a musculoskeletal Digital Assessment Routing Tool (DART): crossover noninferiority randomized pilot trial. JMIR Form Res. Jul 30, 2024;8:e56715. [doi: 10.2196/56715] [Medline: 39078682]

53. Lundberg K, Qin L, Aulin C, van Spil WE, Maurits MP, Knevel R. Population-based user-perceived experience of *Rheumatic?*: a novel digital symptom-checker in rheumatology. RMD Open. Apr 2023;9(2):e002974. [doi: 10.1136/rmdopen-2022-002974] [Medline: 37094982]

54. Martin MJ, Payne KM. Using digital technology and user-centred design to develop a physiotherapy self-referral service for back pain. Physiotherapy. May 2020;107(Suppl 1):e139-e140. [doi: 10.1016/j.physio.2020.03.203]

55. Phillips CJ, Phillips Nee Buck R, Main CJ, et al. The cost effectiveness of NHS physiotherapy support for occupational health (OH) services. BMC Musculoskelet Disord. Feb 23, 2012;13(1):29. [doi: 10.1186/1471-2474-13-29] [Medline: 22361319]

56. Qin L, Zegers F, Selani D, et al. Differentiation of immune mediated versus non immune mediated rheumatic diseases by online symptom checker in real-world patients—multiple diagnoses and particularly fibromyalgia is a stumbling block. Ann Rheum Dis. Jun 2024;83(Suppl 1):2082-2083. [doi: 10.1136/annrheumdis-2024-eular.5438] [Medline: 644868620]

57. Ryan K, Grinbergs P. Demographic analysis of users of a musculoskeletal physiotherapy self-referral digital triage tool in Bromley. Physiotherapy. Jun 2024;123:e210-e211. [doi: 10.1016/j.physio.2024.04.263]

58. Salisbury C, Montgomery AA, Hollinghurst S, et al. Effectiveness of PhysioDirect telephone assessment and advice services for patients with musculoskeletal problems: pragmatic randomised controlled trial. BMJ. Jan 29, 2013;346(7893):f43. [doi: 10.1136/bmj.f43] [Medline: 23360891]

59. Soin A, Hirschbeck M, Verdon M, Manchikanti L. A pilot study implementing a machine learning algorithm to use artificial intelligence to diagnose spinal conditions. Pain Physician. Mar 2022;25(2):171-178. [Medline: 35322974]

60. Tan T, Santosa A, Roslan N, Li J. The development of an AI-based conversational agent for screening of rheumatic diseases [Abstract]. Int J Rheum Dis. 2023;26(9). [doi: 10.1111/1756-185X.14505]

61. Trivedi SV, Batta R, Henao-Romero N, Mondal P, Wilson T, Stempien J. A comparison of self-triage tools to nurse driven triage in the emergency department. PLOS ONE. 2024;19(8):e0297321. [doi: 10.1371/journal.pone.0297321] [Medline: 39196994]

62. Edwards J, Hayden J, Asbridge M, Gregoire B, Magee K. Prevalence of low back pain in emergency settings: a systematic review and meta-analysis. BMC Musculoskelet Disord. Apr 4, 2017;18(1):143. [doi: 10.1186/s12891-017-1511-7] [Medline: 28376873]

63. Lowe C, Atherton L, Lloyd P, Waters A, Morrissey D. Improving safety, efficiency, cost, and satisfaction across a musculoskeletal pathway using the digital assessment routing tool for triage: quality improvement study. J Med Internet Res. Apr 25, 2025;27:e67269. [doi: 10.2196/67269] [Medline: 40279646]

64. Burgess R, Tucker K, Smithson R, Dimbleby P, Casey C. Optimising musculoskeletal patient flow through digital triage and supported self-management: a service evaluation set within community musculoskeletal care. Musculoskelet Care. Dec 2024;22(4):e70013. [doi: 10.1002/msc.70013] [Medline: 39625285]

65. Xu H, Shuttleworth KMJ. Medical artificial intelligence and the black box problem: a view based on the ethical principle of "do no harm". Intell Med. Feb 2024;4(1):52-57. [doi: 10.1016/j.imed.2023.08.001]

66. Mruthyunjaya P, Verma S, Agarwal A, Maharana U, Mandal M, Ahmed S. Right diagnoses but wrong reasoning: current large-language model-based agentic frameworks have flawed clinical reasoning despite high diagnostic accuracy. The Lancet. Preprint posted online on Jul 9, 2025. [doi: 10.2139/ssrn.5339074]

67. Kremer P, Schiebisch H, Lechner F, et al. Comparative analysis of large language models and traditional diagnostic decision support systems for rare rheumatic disease identification. EULAR Rheumatol Open. Jun 2025;1(2):51-59. [doi: 10.1016/j.ero.2025.04.007]

68. Zegers F, Qin L, Selani D, et al. POS1131 prediction models for rheumatic diseases: from clinical simplicity to data-driven complexity with patient-reported symptoms for an online symptom checker. Ann Rheum Dis. Jun 2025;84:1211-1212. [doi: 10.1016/j.ard.2025.06.481] [Medline: 19279015]

69. Segur-Ferrer J, Moltó-Puigmartí C, Pastells-Peiró R, Vivanco-Hidalgo RM. Methodological frameworks and dimensions to be considered in digital health technology assessment: scoping review and thematic analysis. J Med Internet Res. Apr 10, 2024;26:e48694. [doi: 10.2196/48694] [Medline: 38598288]

70. Triage guidelines for orthopaedic optimisation pathway (based on musculoskeletal (MSK) referral) V6.0. South East London Integrated Care Board. 2024. URL: https://www.selondonics.org/wp-content/uploads/MSK-Triage-Guidelines-v6-Final.pdf [Accessed 2025-07-14]

71. Fraenkel L, Bathon JM, England BR, et al. 2021 American College of Rheumatology Guideline for the treatment of rheumatoid arthritis. Arthritis Care Res (Hoboken). Jul 2021;73(7):924-939. [doi: 10.1002/acr.24596] [Medline: 34101387]

72. Adus S, Macklin J, Pinto A. Exploring patient perspectives on how they can and should be engaged in the development of artificial intelligence (AI) applications in health care. BMC Health Serv Res. Oct 26, 2023;23(1):1163. [doi: 10.1186/s12913-023-10098-2] [Medline: 37884940]

73. Lau BHF, Lafave MR, Mohtadi NG, Butterwick DJ. Utilization and cost of a new model of care for managing acute knee injuries: the Calgary Acute Knee Injury Clinic. BMC Health Serv Res. Dec 5, 2012;12(1):445. [doi: 10.1186/1472-6963-12-445] [Medline: 23216946]

74. Lervik LCN, Vasseljen O, Austad B, et al. SupportPrim—a computerized clinical decision support system for stratified care for patients with musculoskeletal pain complaints in general practice: study protocol for a randomized controlled trial. Trials. Apr 11, 2023;24(1):267. [doi: 10.1186/s13063-023-07272-6] [Medline: 37041631]

Abbreviations

AI: artificial intelligence
DART: Digital Assessment Routing Tool
ED: emergency department
LLM: large language model
mHealth: mobile health
OSF: Open Science Framework
PRISMA-S: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Search
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
TRL: technology readiness level