Journal of
Personalized Medicine

MDPI

Review

Artificial Intelligence in Personalized Pain Medicine: Current Evidence, Clinical Applications and Future Directions

Ahmad Mirza [1,*], Anterpreet Dua [2], Paramvir Singh [2] and Zhuo Sun [2]

[1] Department of Medicine, Wellstar MCG Health, Augusta University, Augusta, GA 30912, USA
[2] Department of Anesthesiology, Perioperative and Pain Medicine, Wellstar MCG Health, Augusta University, Augusta, GA 30912, USA
* Correspondence: ahmirza@augusta.edu; Tel.: +1-(706)-721-4761

Abstract

Introduction: Pain is a leading cause of disability because of its subjective, multidimensional and heterogeneous nature. Artificial intelligence (AI), encompassing machine learning (ML), deep learning (DL) and natural language processing (NLP) can integrate multimodal data to support personalized pain care. Personalized pain medicine is defined as the use of patient-specific characteristics to guide individualized assessment, risk stratification and treatment selection. **Methods:** A structured literature review was conducted in PubMed/MEDLINE, Scopus, Web of Science and Google Scholar (January 2020 to 30 June 2026), supplemented by citation searching for foundational studies. Findings were evaluated descriptively because of heterogeneity in populations, algorithms, outcomes and validation methods. A five-stage evidence-maturity framework was applied to classify applications. **Results:** Most published work remains at Stages 1–2 (model development and independent validation). Automated facial-pain assessment can achieve high internal performance (F1 91–96%) that declines markedly with subject-independent validation (F1 40–67%). A postoperative external test set of 120 images, exact 0–10 score prediction reached 53%, with a reported mean error of 2.4 points. A Random Forest model predicted postoperative opioid-requirement categories with 72% accuracy, compared with 71% for multinomial logistic regression; sensitivity was 89% for lower and 43% for higher requirements. The small accuracy difference, incomplete calibration and class-distribution reporting limit claims of incremental value. **Discussion and Conclusions:** AI shows technical promise for automated assessment and risk prediction, but high predictive metrics do not equate to clinical utility. Translation requires external multicenter validation and calibration assessment under clinician supervision and demonstration of improved patient outcomes.

Keywords: artificial intelligence; machine learning; deep learning; pain management; chronic pain; pain assessment; predictive analytics; personalized medicine; personalized care; digital health; large language models

Check for updates

Academic Editor: Ruslan Dorfman

Received: 15 October 2026
Revised: 12 November 2026
Accepted: 17 November 2026
Published: 21 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Pain is one of the most common reasons for seeking medical care and is a multidimensional experience influenced by biological, psychological and social factors. The International Association for the Study of Pain (IASP) defines pain as an unpleasant sensory and emotional experience associated with actual or potential tissue damage; it cannot be adequately characterized by tissue pathology alone [1].

Chronic pain is a particularly complex clinical problem. The IASP recognizes chronic primary pain as pain that persists or recurs for more than 3 months and is associated with significant emotional distress and/or functional disability [2]. This framework supports a biopsychosocial approach in which biological mechanisms, psychological characteristics, behavior, environment and social circumstances contribute to the clinical presentation.

Pain assessment remains centered on patient self-report when reliable self-report is possible. Numeric Rating Scales, Visual Analog Scales and multidimensional instruments remain important components of clinical care. When patients cannot reliably communicate, clinicians may instead use validated observational or behavioral assessment instruments. These approaches are particularly relevant in critical illness, cognitive impairment, infancy and selected neurological or developmental conditions. AI-based approaches are being investigated as an additional source of behavioral or physiological information rather than as a replacement for self-report when self-report is feasible.

The growing availability of electronic health records (EHRs), imaging, physiological signals, wearable sensors, mobile health data and patient-generated health information has created opportunities to integrate heterogeneous clinical information. AI broadly refers to computational methods capable of performing tasks traditionally associated with human intelligence, including pattern recognition, classification and prediction [3,4].

ML methods identify patterns and relationships in structured or semi-structured datasets. DL, a subset of ML, uses multilayer neural networks to learn representations from high-dimensional data such as medical images, physiological signals and videos [3,5]. NLP converts information from clinical encounters, physician notes, nursing documentation and patient-reported descriptions into structured or analyzable data [6].

Computer vision has attracted particular attention in pain medicine because facial movements and other observable behaviors may contain information about pain. Automated facial analysis may therefore provide supplementary information when conventional self-report is unavailable or difficult to obtain [7,8]. However, facial expression is not synonymous with pain. Automated facial-expression systems should therefore be understood as behavioral estimation tools rather than objective measurements of pain itself. Algorithms must account for interindividual differences, movement, lighting, sedation, neurological disease, facial structure, demographic variation and other contextual factors.

The clinical relevance of AI extends beyond assessment. Predictive models may identify patients at increased risk of severe postoperative pain, persistent postsurgical pain, prolonged opioid use, poor rehabilitation outcomes or inadequate treatment response. AI may therefore support a more preventive and individualized approach to care [9–11].

In this review, "personalized medicine," "individualized care" and "precision-based care" are used to denote approaches that explicitly incorporate patient-specific characteristics—pain phenotypes, clinical history, psychological factors, treatment-response patterns. This also includes genetic, pharmacokinetic or physiological data in order to tailor assessment or treatment decisions for the individual. Conventional predictive modeling estimates the probability of an outcome (e.g., high pain score or opioid requirement) for a patient who shares features with a development cohort; it does not, by itself, constitute individualized treatment selection. Genuine personalization requires evidence that an AI-derived recommendation alters management in a manner that improves that patient's outcomes relative to standard care. The present manuscript therefore distinguishes outcome prediction and risk stratification from differential treatment-response prediction, AI-guided therapeutic selection and demonstrated improvement in patient-centered outcomes.

Recent reviews demonstrate the field's rapid expansion. Zhang et al. identified applications of AI in pain assessment, prediction, management and clinical decision support [9]. Antel et al. subsequently identified 17,601 records and included 197 studies on AI in acute

and chronic pain management, underscoring the breadth of the emerging evidence base [10]. Other recent reviews have emphasized the growing role of AI and ML in pain medicine while noting persistent gaps in validation and clinical implementation [11,12].

The specific knowledge gap addressed by this review is the insufficient linkage, in prior studies, between reported computational performance and the clinical requirements of personalized pain medicine. Existing reviews catalog applications and performance metrics. Fewer studies critically evaluate whether those metrics support individualized decision-making, map evidence onto a translational maturity framework or systematically distinguish prediction of average risk from evidence of improved patient-specific outcomes. The present work therefore (1) proposes and applies a five-stage evidence-maturity framework, (2) emphasizes the distinction between technical feasibility and clinical effectiveness and (3) highlights the still-limited evidence that AI-guided treatment selection improves patient-centered outcomes.

Despite these advances, important limitations remain. AI models may be affected by selection bias, incomplete data, demographic underrepresentation, dataset shift, overfitting, limited external validation, class imbalance and inadequate reporting [8,13]. A model that performs well in its development environment may perform substantially worse when applied to a different institution, device, patient population or clinical workflow.

Explainability is another important issue. Complex models may provide predictions without readily understandable reasoning. Although explainable AI seeks to improve interpretability, an explanation generated by a post hoc method should not automatically be taken as evidence that the underlying model is clinically valid or causally informative [14].

Data governance and privacy are also central concerns. AI applications may require access to highly sensitive electronic health records, imaging, physiological, behavioral and patient-generated information. Responsible implementation therefore requires appropriate security, transparency, accountability, equity safeguards and clearly defined data-use practices [15].

Generative AI and large language models (LLMs) have further expanded the potential applications in medicine. These systems may assist with documentation, summarization, information retrieval, patient education, conversational support and extraction of information from clinical narratives. A small, pain-specific study evaluated LLM-based assessment of chronic pain experiences, while other work has examined conversational agents for pain management [16,17]. More broadly, systematic reviews indicate that LLM performance in patient care is heterogeneous and that concerns remain regarding factual accuracy, overconfidence, bias and appropriate human oversight [18].

The objective of this review is to examine current evidence on AI in pain medicine, with emphasis on automated pain assessment and phenotyping, prediction of clinical outcomes, postoperative analgesic planning, personalized care, opioid stewardship, wearable technologies, NLP, generative AI and implementation challenges. Particular attention is given to validation strategy, quantitative performance, generalizability, calibration, clinical utility and the distinction between technical model performance and demonstrated clinical effectiveness.

2. Methods

2.1. Literature Search and Selection

A structured literature search was conducted using PubMed/MEDLINE, Scopus, Web of Science, and Google Scholar. The search was intended to identify representative and methodologically informative studies rather than provide a systematic estimate of the complete evidence base. The first step involved peer-reviewed literature published from January 2020 through 30 June 2026, using terms related to artificial intelligence, machine

learning, deep learning, natural language processing, computer vision, pain assessment, pain management, postoperative pain, opioid use, wearable devices, digital health and precision medicine.

Search strategies were adapted to each database (example PubMed concepts: ("artificial intelligence"[MeSH] OR "machine learning"[MeSH] OR "deep learning" OR "natural language processing" OR "computer vision") AND ("pain"[MeSH] OR "pain management" OR "postoperative pain" OR "chronic pain" OR "opioid")). Inclusion criteria comprised original investigations or focused reviews that reported AI methods applied to pain assessment, phenotyping, prediction, treatment support, opioid stewardship, rehabilitation and wearable device monitoring with extractable information on population, algorithm, validation approach or performance. Exclusion criteria included pure technical algorithm papers without clinical pain relevance, non-English full text when an English abstract provided insufficient methodological detail and opinion pieces lacking primary or secondary data. Studies were classified as "representative" or "methodologically informative" when they illustrated a key application domain, reported quantitative performance with a defined validation strategy or highlighted a critical limitation (e.g., subject-independent validation, external testing or calibration). Formal systematic-review methodology and a PRISMA flow diagram were not required for this review; however, reporting was guided by the Scale for the Assessment of Narrative Review Articles (SANRA) to improve transparency of search, selection and critical appraisal.

2.2. Eligibility and Data Extraction

Studies addressing AI applications in pain assessment, phenotyping, prediction, treatment, opioid stewardship, rehabilitation, wearable monitoring, NLP or clinical decision support were prioritized. Due to substantial heterogeneity in populations, algorithms, outcomes, datasets, and validation methods, findings were summarized descriptively rather than numerically.

Particular attention was given to validation strategy, generalizability, calibration, subgroup performance and clinical relevance. Technical model performance was distinguished from evidence of prospective clinical utility and patient benefit.

Because this was a clinical literature review rather than a formal systematic review, the purpose was to identify representative, methodologically informative evidence in a rapidly evolving field rather than to calculate a pooled effect estimate.

Data extraction also considered the extent to which each application could support personalization at the individual-patient level, including patient-specific phenotyping, treatment-response estimation, risk stratification, longitudinal monitoring and decision support, while distinguishing these potential uses from demonstrated improvements in treatment outcomes. Table 1 provides a taxonomy of major AI applications classified by clinical purpose (assessment, phenotyping, outcome prediction, treatment support, rehabilitation/monitoring).

No formal risk-of-bias assessment with PROBAST + AI was performed for every included study. Instead, the proposed evidence-maturity framework (below) and qualitative appraisal of validation strategy, sample size, external testing and reporting transparency were used to contextualize strength of evidence. The relationship of the framework to existing guidance (TRIPOD+AI, STARD-AI, CONSORT-AI, DECIDE-AI, PROBAST+AI) is described in the Discussion.

Table 1. Taxonomy of AI Applications in Pain Medicine.

Application Domain	AI Approach	Main Clinical Use	Key Limitation/Maturity
Pain assessment	Computer vision/DL	Facial and behavioral pain estimation	Generalizability; demographic bias; Stage 1–2
Pain phenotyping	ML/NLP	Identification of clinically meaningful subgroups	Limited prospective validation of differential response; Stage 1–2
Outcome prediction	ML	Postoperative pain intensity, opioid requirement, prolonged use	Predominantly retrospective; calibration often incomplete; Stage 1–2
Personalized treatment support	ML/DL	Treatment-response prediction and individualized selection	Few prospective trials of AI-directed selection; Stage 1–2
Opioid stewardship	ML	Risk stratification for prolonged use or inadequate analgesia	Cannot replace judgment; risk of stigmatization; Stage 1–2
Rehabilitation & monitoring	ML/DL/ wearables	Continuous activity, sleep, movement, physiological signals	Limited evidence of clinical benefit from triggered interventions; Stage 1–2
NLP & generative AI	NLP/LLMs	Extraction, education, documentation, conversational support	Hallucination, bias, reliability; human oversight required; Stage 1
Clinical decision support	ML/DL/LLMs	Risk prediction and treatment support under supervision	Need for prospective clinical-impact evaluation; Stage 1–2 evidence, with Stage 3 still aspirational

Abbreviations: AI, artificial intelligence; DL, deep learning; LLMs, large language models; ML, machine learning; NLP, natural language processing.

2.3. Proposed Evidence Maturity Framework

For this review, we proposed a five-stage framework to distinguish computational performance from clinical evidence.

Stage 1 represents technical feasibility, including model development and retrospective classification or prediction on a defined dataset.

Stage 2 represents independent validation using patients, institutions, devices or time periods not included during model development.

Stage 3 represents prospective clinical evaluation, including real-time assessment of safety, usability, workflow integration and human factors.

Stage 4 represents clinical-impact evidence from prospective comparative studies demonstrating improvement in patient-centered outcomes, safety, healthcare utilization or cost-effectiveness.

Stage 5 involves post-deployment monitoring of model performance, safety, equity and drift during routine clinical use.

This proposed framework is based on contemporary guidance for reporting, validation, early clinical evaluation, and risk-of-bias assessment, including TRIPOD+AI, STARD-AI, CONSORT-AI, DECIDE-AI and PROBAST+AI [19–23]. It is intended to complement rather than replace those tools: TRIPOD+AI and PROBAST+AI address prediction-model reporting and risk of bias; STARD-AI addresses diagnostic accuracy; CONSORT-AI addresses randomized trials of AI interventions and DECIDE-AI addresses early live clinical evaluation. The five-stage framework supplies a translational ladder that maps the reviewed applications onto the continuum from technical feasibility to routine clinical impact.

3. Results

3.1. Overview of the Evidence

AI applications in pain medicine include automated pain assessment, pain phenotyping, clinical prediction, postoperative analgesic planning, opioid-risk prediction, personalized care, rehabilitation, wearable monitoring, NLP, conversational systems and clinical decision support [9–12].

The evidence remains heterogeneous across populations, data sources, algorithms, outcomes and validation strategies. A 2025 review of ML methods for predicting postoperative pain and opioid use similarly identified substantial heterogeneity in predictor selection, outcome definitions, model development methods and validation strategies [24]. Consequently, reported performance measures should not be interpreted as directly comparable across studies.

The current evidence base is weighted toward Stage 1 and Stage 2 applications: model development and retrospective or independent validation. Evidence from Stage 3 prospective clinical evaluation and Stage 4 clinical-impact studies remains limited. Table 2 places representative applications and studies within the proposed evidence-maturity framework, with brief justification.

Table 2. Evidence-Maturity Framework Applied to Representative AI Applications in Pain Medicine.

Stage	Definition	Representative Applications/Studies	Justification
1	Technical feasibility/model development	Most facial-expression models (internal CV); early LLM pain-narrative studies [16]; wearable feasibility reports [25]; retrospective cohort models [26,27]	Performance shown on development or internal data only
2	Independent/external validation	Fontaine et al. external facial test set [28]; participant-held-out results in Hammal & Cohn [7]	External or participant-independent data were held out from model development; evaluation remains retrospective. Cohort size alone does not establish external validation.
3	Prospective clinical evaluation	Sparse in current pain-AI literature	Real-time safety, usability and workflow data largely absent
4	Clinical-impact evidence	Not yet demonstrated for AI-guided analgesic selection or automated assessment	No comparative trials showing improved patient-centered outcomes attributable to AI
5	Post-deployment monitoring	Not yet applicable at scale in pain medicine	Requires routine clinical use and continuous performance surveillance

The framework complements TRIPOD+AI, STARD-AI, CONSORT-AI, DECIDE-AI and PROBAST+AI by providing an explicit translational ladder; those instruments address reporting, diagnostic accuracy, trial design, early live evaluation and risk-of-bias assessment, respectively. CV, cross-validation.

3.2. Automated Pain Assessment and Phenotyping

Automated pain assessment is among the most extensively studied applications of AI. Computer vision systems can analyze facial movements and other behavioral signals to estimate pain-related outcomes.

Hammal and Cohn demonstrated automated facial pain detection in 25 participants with shoulder injuries [7]. F1-scores ranged from 91–96% with 5-fold cross-validation but dropped to 40–67% with leave-one-subject-out validation. Intraclass correlation also

decreased from 0.85 to 0.55. These findings illustrate the substantial effect of patient-independent validation on apparent model performance.

In a larger postoperative study, Fontaine et al. evaluated 1189 patients and 2810 facial expressions with a ResNet-18 deep-learning system [28]. On an external test set of 120 images, the system predicted the exact 0–10 pain score in 53% of cases, with a reported mean error of 2.4 points. Sensitivity for identifying pain ≥4/10 was 89.7% and for detecting pain ≥7/10 was 77.5%. The external test set of only 120 images is substantially smaller than the development cohort and limits precision of the performance estimates; this limitation is noted in the Abstract and Table 3. Table 3 summarizes key quantitative findings, including sample-size and external-test limitations.

Table 3. Key Quantitative Findings with limitations.

Study	Sample	Pain-Related Outcome	Key Finding and Limitations
Hammal & Cohn [7]	25 participants, shoulder injury	Automated facial pain-intensity detection	F1 91–96% (5-fold CV) declined to 40–67% (leave-one-subject-out); ICC 0.85 → 0.55. Illustrates impact of subject-independent validation.
Amidei et al. [16]	43 written narratives, fibromyalgia	LLM-based assessment of chronic pain experience	Preliminary feasibility only; small exploratory sample does not establish clinical validity.
Fontaine et al. [28]	1189 patients; 2810 expressions; external test $n = 120$ images	Postoperative pain from facial expressions	Exact 0–10 score 53%; reported mean error 2.4 points (the abstract does not identify it as MAE); sensitivity ≥4/10 = 89.7%. External set of 120 images is small; comparison limited to facial-only nurse assessment.
Nair et al. [26]	13,700 ambulatory-surgery patients	Categorized postoperative opioid requirements	Random Forest accuracy ~72% versus 71% for multinomial logistic regression; sensitivity 89% for lower and 43% for higher opioid requirements. Class distribution and calibration incompletely characterized; the small accuracy difference does not establish incremental clinical value.
Saksenberg et al. [27]	2796 elective primary spine-surgery patients	Postoperative pain and opioid consumption	MAE 1.24 pain points; 0.46 MME/kg/day. Retrospective, single-center, population-level models explicitly not intended for individual-level prediction.
Krivicich et al. [29]	Systematic review (orthopedic surgery)	Prolonged postoperative opioid use	Potential discrimination; heterogeneity in populations, predictors, outcomes and validation; external validation largely absent.
Ayena et al. [25]	Scoping review of wearable studies	Chronic pain monitoring and prediction	Feasibility of longitudinal signals shown; evidence of clinical benefit from interventions triggered by wearable data remains limited.

Abbreviations: F1, F1-score; MAE, mean absolute error; MME, morphine milligram equivalents; CV, cross-validation; ICC, intraclass correlation coefficient.

The study's abstract reports a "mean error" of 2.4 points on a 0–10 scale but does not identify it as mean absolute error (MAE); the manuscript therefore retains the reported terminology. Whether 53% exact-score accuracy or a 2.4-point mean error is sufficient to

support clinical decision-making depends on the intended use (e.g., screening for moderate–severe pain versus precise titration). Comparison with conventional nurse assessment limited to facial expression is informative for technical feasibility but does not establish superiority over comprehensive clinical assessment that incorporates self-report, vital signs and context [28].

These results support the feasibility of automated facial-expression–based estimation of postoperative pain, but they should not be interpreted as evidence that facial analysis constitutes an objective replacement for patient-reported pain. When reliable self-report is available, automated assessment should therefore be considered supplementary.

The broader literature similarly indicates that automated facial-pain recognition can achieve high performance under controlled or internal validation conditions but may show reduced generalizability across individuals and clinical contexts [8].

AI-based pain phenotyping is a related but distinct application. Rather than estimating momentary pain intensity, phenotyping approaches aim to identify clinically meaningful subgroups or patterns based on symptoms, clinical history, imaging, physiological signals or treatment response. These approaches may eventually contribute to mechanism-informed care, but phenotype discovery must be distinguished from demonstrating that an AI-derived subgroup responds differently to a specific intervention.

3.3. Multimodal Clinical Data Analysis and Prediction Modelling

AI can integrate clinical history, imaging, physiological measurements, laboratory findings, EHR data, and patient-generated information [3,4,9,11]. DL is particularly suited to high-dimensional imaging data, whereas NLP can extract clinically relevant information from subjective documentation [5,6].

Multimodal approaches may offer an opportunity to integrate complementary biological, psychological, functional and contextual information. However, whether multimodal models provide greater clinical value than simpler models remains to be established. Multimodal models also increase requirements for data harmonization, missing-data management, interoperability, privacy protection and external validation.

A further concern is that multimodal models may inadvertently learn site- or device-specific characteristics rather than clinically meaningful relationships. External validation should therefore assess not only whether performance is preserved but also whether model inputs remain available, consistently measured and clinically interpretable in the target setting. Table 3 summarizes the key quantitative findings from these representative studies.

3.4. Postoperative Analgesic Planning

Machine-learning approaches have been evaluated for predicting postoperative pain intensity and opioid requirements using perioperative clinical variables. Nair et al. analyzed 13,700 ambulatory-surgery patients. The Random Forest model achieved 72% accuracy for categorizing postoperative opioid requirements, compared with 71% for multinomial logistic regression [26]. Sensitivity for the Random Forest model was 89% for lower opioid requirements but only 43% for higher requirements [26]. The one-percentage-point accuracy difference does not establish a meaningful advantage over conventional regression. Class distribution and calibration were not fully characterized, limiting interpretation of clinical value.

Langford et al. reviewed contemporary machine-learning methods for postoperative pain and opioid-use prediction and identified substantial variation in predictor selection, outcome definitions, modeling strategies and validation procedures across studies [24]. In elective primary spine surgery, Saksenberg et al. evaluated 2796 patients using population-level machine-learning models to estimate postoperative pain and opioid consumption [27].

The reported mean absolute error was approximately 1.24 points for postoperative pain and 0.46 MME/kg/day for opioid consumption [27]. The authors explicitly state that these models were intended to predict population-level trends and were not intended for individual-level prediction. These findings demonstrate that perioperative variables can be used to generate quantitative estimates for pain intensity and opioid exposure at the cohort level. The available studies evaluated outcomes using different time points and definitions, limiting direct comparison of predictive performance between models [24,27]. Most reported evidence has been derived from retrospective datasets rather than prospective implementation of algorithm-guided analgesic plans [24,27]. Thus, the current results demonstrate predictive capability but provide limited evidence that incorporating these predictions into prescribing decisions changes postoperative outcomes.

3.5. Opioid-Risk Prediction

Machine-learning models have also been investigated for identifying patients at increased risk of prolonged postoperative opioid use. Krivicich et al. systematically reviewed 10 studies applying machine-learning algorithms to prolonged postoperative opioid use after orthopedic surgery. The reviewed models demonstrated potentially useful discrimination, but study populations, predictors, outcome definitions and modeling methods were heterogeneous and external validation was not reported across the included studies [29]. The outcome definitions for prolonged opioid use also varied between investigations, contributing to differences in reported model performance [29]. Nair et al. additionally demonstrated the feasibility of predicting categorized postoperative opioid requirements in a large ambulatory-surgery population [26]. The postoperative prediction literature therefore includes both estimation of short-term opioid requirements and identification of longer-term opioid-use risk [24,26,29]. These applications address related but distinct outcomes, with opioid requirement reflecting immediate analgesic needs and prolonged use representing a longer-term medication trajectory. Collectively, the available results support the feasibility of algorithmic opioid-risk stratification, although variation in populations, predictors, endpoints and validation methods limits direct comparison between individual models.

Importantly, prediction of opioid requirement or prolonged use does not equate to demonstrable improvement in opioid stewardship. Models must be evaluated for calibration, clinically relevant decision thresholds, potential stigmatization of high-risk labels and whether their use under clinician supervision actually reduces inappropriate exposure or improves analgesia without increasing harm.

3.6. Personalized Care and Rehabilitation

AI-based personalized care applications have focused on integrating multiple clinical characteristics to support individualized assessment and treatment selection. Reviews of AI in pain medicine describe potential use of machine learning for treatment-response prediction, patient phenotyping, medication selection, interventional planning and rehabilitation support [9–12]. Multimodal approaches can incorporate clinical history, imaging, physiological measurements, functional information and treatment-response data within predictive models [9,11,12]. These approaches provide a framework for identifying clinically relevant patient subgroups that may differ in symptom patterns or treatment response. AI-based models may also incorporate longitudinal information that is difficult to synthesize manually across large clinical datasets [11,12]. Current evidence includes exploratory prediction of treatment response and clinical characteristics rather than large prospective trials demonstrating improved outcomes from AI-directed treatment selection [9–12]. Rehabilitation-related applications similarly have the potential to combine functional mea-

sures, activity patterns, clinical information and longitudinal treatment data. The available literature describes potential applications for computationally assisted personalization and rehabilitation monitoring, but evidence that algorithm-guided treatment selection improves functional recovery remains limited.

Pharmacological determinants of individualized analgesia—pain mechanisms, pharmacokinetic and pharmacodynamic variability, pharmacogenetic characteristics, comorbidities, concomitant medications and prior treatment responses—remain under-represented in most current AI models. Integration of these variables into predictive systems is theoretically attractive for mechanism-based selection and adverse-reaction avoidance, yet clinical evidence supporting such integration is still largely investigational. Distinguishing applications supported by the empirical data from those that remain conceptual is essential for realistic translational claims.

3.7. Wearable Technologies

Wearable technologies provide continuous or repeated measurements that can supplement conventional clinic-based pain assessments. Ayena et al. reviewed wearable-device applications for chronic pain prediction and identified activity, movement, physiological and behavioral signals as potentially relevant sources of longitudinal information [25]. The reviewed technologies included sensors capable of capturing information associated with physical activity and physiological state over extended periods [25]. Such data can provide repeated measurements between clinical encounters rather than relying exclusively on isolated assessments obtained during visits [25]. Wearable-derived information may therefore contribute to characterization of temporal changes in pain-related behavior and function. The review also identified substantial variation in sensor characteristics, data-processing approaches, security practices and standards compliance across studies [25]. Evidence for prospective clinical utility and demonstrated patient benefit was less developed than evidence showing the feasibility of collecting and analyzing wearable-derived signals [25]. Device adherence, sensor placement, environmental influences, interoperability and data standardization were recurring methodological considerations [25]. Overall, the results indicate that wearable technologies can generate clinically relevant longitudinal datasets potentially relevant to chronic pain monitoring and prediction, but their effectiveness as components of interventions remains insufficiently established.

4. Discussion

4.1. Principal Findings

This review indicates that AI has evolved from predominantly experimental applications to increasingly sophisticated tools for pain assessment, prediction, phenotyping, treatment support, opioid-risk stratification, wearable monitoring, NLP and clinical decision support.

The strongest evidence currently focuses on technical feasibility and retrospective or independent model performance rather than definitive improvement in patient outcomes.

The studies above indicate that AI may provide useful complementary information but should not be considered interchangeable with patient-reported pain assessment.

A clearer relationship must be established between AI-driven prediction, personalized treatment selection, treatment-response monitoring and measurable improvements in patient outcomes. Predicting pain scores or opioid requirements demonstrates technical feasibility (Stages 1–2) but does not, by itself, establish that the prediction alters clinician behavior in a way that improves analgesia, function, safety or equity for the individual patient (Stages 3–4).

4.2. Generalizability and Dataset Shift

One of the most important findings in the literature is that internal performance should not be equated with clinical validity. The decline in performance observed with patient-independent validation in automated facial pain assessment demonstrates that models may learn participant-specific characteristics [7]. Similar problems may arise when models capture institution-specific documentation practices, imaging characteristics, device-related artifacts, clinician behavior or population characteristics.

Dataset shift can occur when the distributions of predictors, outcomes or predictor-outcome relationships change between development and deployment. Even modest differences between training and clinical populations can affect performance.

External validation should ideally include patients from different institutions, geographic regions, demographic groups, devices and time periods. Temporal validation can also assess whether model performance remains stable as clinical practice evolves.

4.3. From Technical Performance to Clinical Utility

AI studies report a range of metrics, including accuracy, sensitivity, specificity, F1-score, intraclass correlation, mean absolute error, area under the receiver operating characteristic curve and other discrimination measures. These metrics capture different aspects of performance and should be selected according to the intended clinical application.

For continuous outcomes such as pain intensity, mean absolute error provides an intuitive estimate of prediction error on the original measurement scale. For risk prediction, discrimination should be accompanied by calibration assessment. A model may distinguish higher-risk from lower-risk patients while systematically overestimating or underestimating absolute risk.

Clinical utility is a step beyond discrimination and calibration. Decision-curve analysis and related approaches can help determine whether using a prediction model provides net benefit across clinically relevant decision thresholds. Ultimately, a prediction is useful only if it can appropriately change clinical management and if that change improves outcomes or reduces harm.

The reported 53% exact pain-score prediction and mean error of 2.4 points, together with the 72% accuracy for Random Forest opioid-requirement categorization versus 71% for multinomial logistic regression, must be interpreted in light of the small external facial-image test set, class-specific sensitivity, calibration, class distribution and the clinical consequences of misclassification [26,28]. Without such contextualization, performance metrics risk overstating readiness for clinical decision support.

The quantitative findings summarized in Table 3 should therefore be interpreted as measures of study-specific model performance rather than as evidence of equivalent clinical effectiveness across applications.

4.4. Postoperative Pain and Opioid Prediction

Postoperative prediction is a practical application because analgesic planning routinely uses information available before or immediately after surgery.

The study by Nair et al. demonstrated the feasibility of predicting categorized postoperative opioid requirements using perioperative information from 13,700 ambulatory-surgery patients [26]. The Random Forest model achieved approximately 72% accuracy, compared with 71% for multinomial logistic regression. Sensitivity was 89% for lower opioid requirements and 43% for higher requirements [26]. This small difference in overall accuracy, together with the lower sensitivity for higher requirements, illustrates why accuracy alone does not fully characterize model performance. Class distribution, calibration and clinically relevant decision thresholds should also be considered.

The systematic review by Senders et al. provides broader evidence that ML can predict heterogeneous neurosurgical outcomes, but these results should not be taken as evidence of equivalent performance in pain-specific prediction [30].

Saksenberg et al. subsequently demonstrated ML-based population-level prediction of postoperative pain and opioid consumption in 2796 elective primary spine-surgery cases [27]. The reported MAEs of approximately 1.24 points for postoperative pain and 0.46 MME/kg/day for opioid consumption demonstrate predictive feasibility. However, the retrospective design, single-center setting and population-level nature of the analysis limit conclusions about individual treatment decisions.

Prospective studies are needed to determine whether AI-guided decisions reduce uncontrolled pain, excessive opioid exposure, adverse events or healthcare utilization.

4.5. Personalized Care

The heterogeneity of pain makes precision treatment particularly appealing. Patients with similar diagnostic labels may differ substantially in pain mechanisms, psychological characteristics, functional goals, medication tolerance and treatment response.

AI could integrate biological, psychological, functional, imaging and treatment-response variables to identify individualized therapeutic strategies [11,12]. Potential applications include medication selection, interventional treatment planning, rehabilitation and prediction of treatment response.

However, a model that predicts treatment response does not automatically establish that treatment should be individualized based on that prediction. The clinically important question is whether an AI-guided strategy yields better outcomes than standard clinical decision-making.

Future trials should therefore compare AI-assisted treatment selection with usual care and assess pain intensity, physical function, quality of life, adverse events, treatment adherence, patient satisfaction and cost-effectiveness.

The potential role of AI in opioid stewardship is clinically important yet ethically sensitive. Predicting prolonged opioid use could help identify patients who may benefit from additional counseling, multimodal analgesia, closer follow-up, behavioral support or rehabilitation [29].

However, risk prediction must not replace individualized assessment. Patients identified as high risk should not automatically be denied appropriate analgesia. Conversely, patients classified as low risk should not be assumed to be at no risk of opioid-related adverse outcomes.

AI should therefore be integrated into broader opioid stewardship programs, with explicit clinical oversight, monitoring for unintended consequences and ongoing evaluation of subgroup performance.

Translation of AI-generated predictions into actionable clinical decisions requires clinician supervision at every step: the model output is an additional data point, not an autonomous order. Decision-support interfaces should present uncertainty, key contributing features (when reliable) and explicit recommendations for human review rather than automated prescribing.

4.6. Wearables and Continuous Monitoring

Wearable technologies may transform pain assessment from episodic measurement to longitudinal monitoring.

Traditional clinical assessment captures pain at discrete time points, whereas wearable devices can continuously capture activity, sleep, movement and physiological variables [25].

This information may be particularly useful in chronic pain, where fluctuations and behavioral changes may be more informative than a single clinic-based pain score.

However, wearable systems pose challenges related to device adherence, sensor placement, environmental conditions, data quality, interoperability and standardization. Future studies should determine whether wearable-derived information can trigger interventions that improve clinically meaningful outcomes rather than merely demonstrating correlations with pain.

The key clinical requirements for translating these technologies into practice are summarized in Table 4.

Table 4. Key Requirements for Clinical AI in Pain Medicine.

Domain	Requirement
Validation	External, preferably multicenter validation; temporal validation when feasible
Performance	Discrimination, calibration, uncertainty quantification, and systematic subgroup analysis
Clinical utility	Prospective assessment of decision impact and clinically meaningful patient outcomes under clinician supervision
Safety	Human oversight, failure-mode assessment, drift detection, and monitoring for unintended consequences (e.g., opioid restriction)
Equity	Representative data and systematic evaluation across age, sex, race, clinical condition and communication ability
Governance	Privacy, security, transparency, accountability, and appropriate data-use practices
Reproducibility	Transparent reporting of data processing, model development, validation strategy and performance metrics
Implementation	Workflow integration, usability, clinician workload, alert burden and patient acceptance

4.7. NLP and Generative AI

NLP can extract clinically important information from large volumes of unstructured medical documentation [6]. This may improve phenotyping, longitudinal characterization, cohort identification and clinical summarization.

LLMs and conversational systems may also assist with patient education, documentation, information retrieval, and self-management support [16–18]. However, the emerging evidence base does not yet support autonomous clinical decision-making. A 2025 review of conversational agents in pain management found a limited evidence base and emerging analyses have raised concerns about variability and potential demographic disparities in LLM outputs [17,31].

Generative AI therefore requires a human-in-the-loop model in which clinicians remain responsible for interpretation and decision-making. Systems should be evaluated for factual accuracy, completeness, overconfidence, bias, reproducibility and failure modes in clinically realistic scenarios.

4.8. Ethical and Equity Considerations

AI can reproduce or amplify inequities present in its training data. If certain demographic groups are underrepresented, models may perform less accurately for those groups.

This issue is especially important for automated facial pain assessment. Facial analysis systems should be evaluated across diverse demographic groups, including age, skin tone, facial structure and clinical conditions.

Algorithmic bias is an important concern in pain-related applications, particularly when training data underrepresent demographic or clinical groups. The literature has raised concerns about bias in AI-based pain management and language-model outputs [31]. However, the extent to which existing automated facial-pain systems perform differently across demographic groups and clinical conditions remains insufficiently characterized. Opioid-risk prediction models carry additional risks of stigmatization, inappropriate risk classification and unjustified restriction of analgesia; these consequences should be prospectively monitored.

Fairness assessment should extend beyond aggregate performance. Relevant analyses may include subgroup sensitivity, specificity, calibration, false-positive and false-negative rates, and differences in downstream recommendations.

Privacy is equally important. AI applications that integrate EHRs, wearable data, imaging, and patient-generated information create large repositories of sensitive health data. Appropriate governance should address data minimization, access control, encryption, consent, secondary use, data retention, cybersecurity and accountability [15].

4.9. Explainability and Clinical Trust

Explainability may improve clinicians' understanding of AI recommendations, but it should not be treated as evidence of model validity.

Ghassemi et al. cautioned that some explainability approaches may produce explanations that do not necessarily align with the actual computational basis of model predictions [14].

The clinically useful objective should therefore be interpretable and testable decision support. Clinicians need to know which patient characteristics contributed to a prediction, how reliable the prediction is, how to interpret uncertainty and whether similar patients were represented in the validation population.

4.10. Reporting, Reproducibility, and Evaluation Standards

Contemporary AI research should follow guidance appropriate to the study design. TRIPOD+AI provides updated guidance for reporting clinical prediction models developed using regression or ML methods [19]. STARD-AI provides guidance for diagnostic-accuracy studies involving AI [20]. CONSORT-AI extends CONSORT reporting for randomized trials involving AI interventions [21]. DECIDE-AI addresses early-stage live clinical evaluation of AI-based decision-support systems, with an emphasis on clinical performance, safety, human factors and preparation for larger trials [22]. PROBAST+AI provides an updated framework for assessing the risk of bias and applicability of prediction models developed using regression or AI methods [23].

These documents serve complementary purposes rather than serving as interchangeable frameworks. Together, they provide a useful structure for transparent reporting, critical appraisal, clinical evaluation and assessment of risk of bias.

A clinically mature AI system for chronic pain management should progress through several stages:

Model development → internal validation → independent external validation → prospective clinical evaluation → comparative clinical-impact assessment → ongoing post-deployment monitoring.

This staged approach is intended to reduce the risk of deploying models that perform well in development but poorly in real-world practice.

Reproducibility also requires reporting patient-selection criteria, missing-data handling, pre-processing, feature selection, model architecture, hyperparameter tuning, validation strategy, performance metrics, confidence intervals where appropriate, calibration, subgroup analyses, and relevant code or data-access information when feasible.

4.11. Clinical Implementation and Model Drift

Clinical implementation introduces problems that are not apparent in retrospective datasets. Workflow integration may alter how clinicians interact with the system, which in turn can affect model use and patient outcomes.

Implementation studies should therefore evaluate clinician workload, alert burden, usability, interpretability, patient acceptance, workflow disruption and unintended consequences. Early clinical evaluation should address both the AI system and the human-computer interaction it entails [22].

Because patient populations, treatment patterns, clinical documentation, devices, and data-generating processes can change over time, deployed models should undergo ongoing monitoring for performance drift. A model should not be considered permanently validated merely because it performed adequately in its original development cohort.

Regulatory frameworks for AI-based clinical software (FDA guidance on software as a medical device and applicable European regulations) should be distinguished from scientific reporting guidelines; both are relevant to eventual clinical deployment but address different questions.

4.12. Limitations of the Review

This review is descriptive rather than systematic and therefore does not provide pooled effect estimates or a formal quantitative synthesis. The search was designed to identify representative, methodologically informative evidence across a rapidly evolving field rather than to estimate a single summary effect.

Substantial heterogeneity existed across the included studies in populations, pain conditions, datasets, algorithms, outcomes, class distributions and validation strategies. Consequently, reported performance measures cannot be directly compared across studies without accounting for these methodological differences.

Foundational studies published before 2020 were retained if they provided directly relevant methodological or conceptual evidence. The reference list should therefore not be interpreted as limited to the 2020–2026 period.

Because the field is evolving rapidly, the literature search cutoff of 30 June 2026 means that publications after that date were not considered. This is particularly relevant to generative AI and LLM research, where evidence and model capabilities are changing quickly.

Finally, much of the available evidence remains retrospective or focused on model development rather than on prospective patient outcomes. Technical performance is therefore demonstrated more extensively than clinical effectiveness.

Additional limitations include the absence of a formal PROBAST+AI risk-of-bias appraisal for each primary study and limited coverage of pediatric and communication-impaired populations. The review also provides limited depth on physiological and neurophysiological approaches. It identified a lack of multimodal datasets beyond the cited examples, limited coverage of interventional pain and neuromodulation applications, and sparse primary literature on wearables, NLP and generative AI within the pain domain. Dataset imbalance, potential overfitting, data leakage, explainability gaps, demographic bias and the lack of prospective clinical-impact evidence are recurrent methodological concerns that constrain claims of readiness for routine personalized care.

5. Future Directions

Future research should prioritize multicenter datasets, standardized definitions of pain and treatment outcomes, independent external validation, calibration assessment and prospective clinical evaluation.

Studies should assess whether AI-assisted decision-making improves clinically meaningful outcomes, including pain intensity, physical function, quality of life, patient satisfaction, medication safety, opioid exposure, rehabilitation adherence, healthcare utilization and cost-effectiveness.

Prediction studies should report discrimination and calibration and evaluate clinically relevant decision thresholds. Decision-curve analysis or other clinical-utility approaches should be considered when appropriate.

Subgroup performance should be evaluated to identify potential disparities across demographic and clinical populations. For facial-analysis systems, this should include assessments across relevant demographic characteristics and clinical conditions. For LLMs and conversational systems, evaluation should also consider demographic variation in recommendations and language.

Prospective implementation studies should examine workflow integration, clinician workload, alert burden, usability, patient acceptance and cost. Reporting should align with contemporary frameworks such as TRIPOD+ AI, STARD-AI, CONSORT-AI, and DECIDE-AI, tailored to study design [19–21].

Because patient populations, treatment patterns, and data-generating processes may change over time, deployed models should undergo continuous performance monitoring to detect drift and unexpected harms.

Generative AI requires additional safeguards. Future studies should determine when LLMs can safely support documentation, patient education, clinical summarization, information retrieval and pain self-management without introducing unacceptable risks from inaccurate, unsupported or biased information.

Most importantly, future research should move beyond demonstrating that AI can predict an outcome and determine whether using that prediction alters clinical decisions in ways that improve patient outcomes.

6. Conclusions

Artificial intelligence is increasingly relevant to modern pain medicine, with applications in automated pain assessment, pain phenotyping, clinical prediction and postoperative analgesic planning. These applications also include opioid-risk stratification, personalized care, wearable monitoring, NLP, conversational systems and clinical decision support.

Quantitative evidence demonstrates promising technical performance but also highlights important limitations.

These studies demonstrate both the potential and the limitations of AI. High algorithmic performance does not automatically establish generalizability, clinical utility or improved patient outcomes. Predicting treatment response does not, by itself, constitute personalized treatment, and automated behavioral assessment does not constitute an objective replacement for patient-reported pain when reliable self-report is available.

AI should therefore be implemented as a clinical decision-support technology that augments rather than replaces healthcare professionals. Successful adoption will require multicenter external validation, representative datasets, standardized outcomes and transparent reporting.

With rigorous validation and responsible implementation, AI may eventually support more precise, preventive and individualized pain management. The ultimate measure of

progress should be whether AI improves the safety, effectiveness, equity and experience of pain care rather than merely producing better computational metrics.

Author Contributions: A.M.: Conceptualization, methodology, writing—original draft, critical revision and manuscript finalization. A.D.: Literature search, referencing, manuscript editing and writing—review and editing. P.S.: Writing—review and editing. Z.S.: Literature search and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

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

References

1. Raja, S.N.; Carr, D.B.; Cohen, M.; Finnerup, N.B.; Flor, H.; Gibson, S.; Keefe, F.J.; Mogil, J.S.; Ringkamp, M.; Sluka, K.A.; et al. The revised International Association for the Study of Pain definition of pain: Concepts, challenges, and compromises. *Pain* **2020**, *161*, 1976–1982. [CrossRef] [PubMed]
2. Nicholas, M.; Vlaeyen, J.W.S.; Rief, W.; Barke, A.; Aziz, Q.; Benoliel, R.; Cohen, M.; Evers, S.; Giamberardino, M.A.; Goebel, A.; et al. The IASP classification of chronic pain for ICD-11: Chronic primary pain. *Pain* **2019**, *160*, 28–37. [CrossRef] [PubMed]
3. Topol, E.J. High-performance medicine: The convergence of human and artificial intelligence. *Nat. Med.* **2019**, *25*, 44–56. [CrossRef] [PubMed]
4. Davenport, T.; Kalakota, R. The potential for artificial intelligence in healthcare. *Future Healthc. J.* **2019**, *6*, 94–98. [CrossRef] [PubMed]
5. Lundervold, A.S.; Lundervold, A. An overview of deep learning in medical imaging focusing on MRI. *Z. Med. Phys.* **2019**, *29*, 102–127. [CrossRef] [PubMed]
6. Kreimeyer, K.; Foster, M.; Pandey, A.; Arya, N.; Halford, G.; Jones, S.F.; Forshee, R.; Walderhaug, M.; Botsis, T. Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review. *J. Biomed. Inform.* **2017**, *73*, 14–29. [CrossRef] [PubMed]
7. Hammal, Z.; Cohn, J.F. Automatic detection of pain intensity. In *Proceedings of the International Conference on Multimodal Interaction*; Association for Computing Machinery: New York, NY, USA, 2012; Volume 2012, pp. 47–52. [CrossRef] [PubMed]
8. Hassan, T.; Seus, D.; Wollenberg, J.; Weitz, K.; Kunz, M.; Lautenbacher, S.; Garbas, J.-U.; Schmid, U. Automatic detection of pain from facial expressions: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 1815–1831. [CrossRef] [PubMed]
9. Zhang, M.; Zhu, L.; Lin, S.Y.; Herr, K.; Chi, C.L.; Demir, I.; Lopez, K.D.; Chi, N.-C. Using artificial intelligence to improve pain assessment and pain management: A scoping review. *J. Am. Med. Inform. Assoc.* **2023**, *30*, 570–587. [CrossRef] [PubMed]
10. Antel, R.; Whitelaw, S.; Gore, G.; Ingelmo, P. Moving towards the use of artificial intelligence in pain management. *Eur. J. Pain* **2025**, *29*, e4748. [CrossRef] [PubMed]
11. Hagedorn, J.M.; George, T.K.; Aiyer, R.; Schmidt, K.; Halamka, J.; D'Souza, R.S. Artificial intelligence and pain medicine: An introduction. *J. Pain Res.* **2024**, *17*, 509–518. [CrossRef] [PubMed]
12. Abd-Elsayed, A.; Robinson, C.L.; Marshall, Z.; Diwan, S.; Peters, T. Applications of artificial intelligence in pain medicine. *Curr. Pain Headache Rep.* **2024**, *28*, 229–238. [CrossRef] [PubMed]
13. Beam, A.L.; Manrai, A.K.; Ghassemi, M. Challenges to the reproducibility of machine learning models in health care. *JAMA* **2020**, *323*, 305–306. [CrossRef] [PubMed]
14. Ghassemi, M.; Oakden-Rayner, L.; Beam, A.L. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit. Health* **2021**, *3*, e745–e750. [CrossRef] [PubMed]
15. World Health Organization. *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*; World Health Organization: Geneva, Switzerland, 2021.
16. Amidei, J.; Nieto, R.; Kaltenbrunner, A.; Ferreira De Sá, J.G.; Serrat, M.; Albajes, K. Exploring the capacity of large language models to assess the chronic pain experience: Algorithm development and validation. *J. Med. Internet Res.* **2025**, *27*, e65903. [CrossRef] [PubMed]

17. Souza, F.L.; Bowman, H.; Yang, F.; Hesam-Shariati, N.; Linke, J.; Gilanyi, Y.L.; Jones, M.D.; Z-Pinto, R.; McAuley, J.H.; Rizzo, R.R.N. Conversational agents to support pain management: A scoping review. *Eur. J. Pain* **2025**, *29*, e70016. [CrossRef] [PubMed]
18. Busch, F.; Hoffmann, L.; Rueger, C.; van Dijk, E.H.C.; Kader, R.; Ortiz-Prado, E.; Makowski, M.R.; Saba, L.; Hadamitzky, M.; Kather, J.N.; et al. Current applications and challenges in large language models for patient care: A systematic review. *Commun. Med.* **2025**, *5*, 26. [CrossRef] [PubMed]
19. Dijkstra, P.; Greenhalgh, T.; Mekki, Y.M.; Morley, J. How to read a paper involving artificial intelligence (AI). *BMJ Med.* **2025**, *4*, e001394. [CrossRef]
20. Williams, S.C.; Khan, D.Z.; Bano, S.; Blandford, A.; Clarkson, M.J.; Costanza, E.; Mazomenos, E.B.; Stoyanov, D.; McCulloch, P.; Marcus, H.J. A surgical approach to building impactful artificial intelligence. *npj Digit. Med.* **2025**, *8*, 709. [CrossRef] [PubMed]
21. Liu, X.; Rivera, S.C.; Moher, D.; Calvert, M.J.; Denniston, A.K.; Ashrafian, H.; Beam, A.L.; Chan, A.-W.; Collins, G.S.; Darzi, A.; et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. *BMJ* **2020**, *370*, m3164. [CrossRef] [PubMed]
22. Vasey, B.; Nagendran, M.; Campbell, B.; Clifton, D.A.; Collins, G.S.; Denaxas, S.; Denniston, A.K.; Faes, L.; Geerts, B.; Ibrahim, M.; et al. Publisher Correction: Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat. Med.* **2022**, *28*, 2218. [CrossRef] [PubMed]
23. Moons, K.G.M.; Damen, J.A.A.; Kaul, T.; Hooft, L.; Andaur Navarro, C.; Dhiman, P.; Beam, A.L.; Van Calster, B.; Celi, L.A.; Denaxas, S.; et al. PROBAST+AI: An updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ* **2025**, *388*, e082505. [CrossRef] [PubMed]
24. Langford, D.J.; Reichel, J.F.; Zhong, H.; Basseri, B.H.; Koch, M.P.; Kolady, R.; Liu, J.; Sideris, A.; Dworkin, R.H.; Poeran, J.; et al. Machine learning research methods to predict postoperative pain and opioid use: A literature review. *Reg. Anesth. Pain Med.* **2025**, *50*, 102–109. [CrossRef] [PubMed]
25. Ayena, J.C.; Bouayed, A.; Ben Arous, M.; Ouakrim, Y.; Loulou, K.; Ameyed, D.; Savard, I.; El Kamel, L.; Mezghani, N. Predicting chronic pain using wearable devices: A scoping review of sensor capabilities, data security, and standards compliance. *Front. Digit. Health* **2025**, *7*, 1581285. [CrossRef] [PubMed]
26. Nair, A.A.; Velagapudi, M.A.; Lang, J.A.; Behara, L.; Venigandla, R.; Velagapudi, N.; Fong, C.T.; Horibe, M.; Lang, J.D.; Nair, B.G. Machine learning approach to predict postoperative opioid requirements in ambulatory surgery patients. *PLoS ONE* **2020**, *15*, e0236833. [CrossRef] [PubMed]
27. Saksenberg, D.L.; Lee, R.Y.; Vaidya, S.; Yuan, A.; Jeong, S.J.; Lockwood, N.; Li, J.; Russo, G.S. Machine learning approach to predict postoperative pain and opioid usage in elective primary spine surgery: A retrospective study. *J. Pain Res.* **2026**, *19*, 596384. [CrossRef] [PubMed]
28. Fontaine, D.; Vielzeuf, V.; Genestier, P.; Limeux, P.; Santucci-Sivilotto, S.; Mory, E.; Darmon, N.; Lanteri-Minet, M.; Mokhtar, M.; Laine, M.; et al. Artificial intelligence to evaluate postoperative pain based on facial expression recognition. *Eur. J. Pain* **2022**, *26*, 1282–1291. [CrossRef] [PubMed]
29. Krivicich, L.M.; Jan, K.; Kunze, K.N.; Rice, M.; Nho, S.J. Machine learning algorithms can be reliably leveraged to identify patients at high risk of prolonged postoperative opioid use following orthopedic surgery: A systematic review. *HSS J.* **2024**, *20*, 589–599. [CrossRef] [PubMed]
30. Senders, J.T.; Staples, P.C.; Karhade, A.V.; Zaki, M.M.; Gormley, W.B.; Broekman, M.L.D.; Smith, T.R.; Arnaout, O. Machine learning and neurosurgical outcome prediction: A systematic review. *World Neurosurg.* **2018**, *109*, 476–486.e1. [CrossRef] [PubMed]
31. Young, C.C.; Enichen, E.; Rao, A.; Succi, M.D. Racial, ethnic, and sex bias in large language model opioid recommendations for pain management. *Pain* **2025**, *166*, 511–517. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.