

Artificial Intelligence and Machine Learning in Cancer Related Pain: A Systematic Review

Vivian Salama¹, Brandon Godinich^{1,2}, Yimin Geng³, Laia Humbert-Vidan¹, Laura Maule¹, Kareem A. Wahid¹, Mohamed A. Naser¹, Renjie He¹, Abdallah S.R. Mohamed¹, Clifton D. Fuller¹ and Amy C. Moreno¹.

1 Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

2 Department of Medical Education, Paul L. Foster School of Medicine, Texas Tech Health Sciences Center, El Paso, TX, USA.

3 Research Medical Library, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

Abstract:

Background/objective: Pain is a challenging multifaceted symptom reported by most cancer patients, resulting in a substantial burden on both patients and healthcare systems. This systematic review aims to explore applications of artificial intelligence/machine learning (AI/ML) in predicting pain-related outcomes and supporting decision-making processes in pain management in cancer.

Methods: A comprehensive search of Ovid MEDLINE, EMBASE and Web of Science databases was conducted using terms including “Cancer”, “Pain”, “Pain Management”, “Analgesics”, “Opioids”, “Artificial Intelligence”, “Machine Learning”, “Deep Learning”, and “Neural Networks” published up to September 7, 2023. The screening process was performed using the Covidence screening tool. Only original studies conducted in human cohorts were included. AI/ML models, their validation and performance and adherence to TRIPOD guidelines were summarized from the final included studies.

Results: This systematic review included 44 studies from 2006-2023. Most studies were prospective and uni-institutional. There was an increase in the trend of AI/ML studies in cancer pain in the last 4 years. Nineteen studies used AI/ML for classifying cancer patients’ pain development after cancer therapy, with median AUC 0.80 (range 0.76-0.94). Eighteen studies focused on cancer pain research with median AUC 0.86 (range 0.50-0.99), and 7 focused on applying AI/ML for cancer pain management decisions with median AUC 0.71 (range 0.47-0.89). Multiple ML models were investigated with median AUC across all models in all studies (0.77). Random forest models demonstrated the highest performance (median AUC 0.81), lasso models had the highest median sensitivity (1), while Support Vector Machine had the highest median specificity (0.74). Overall adherence of included studies to TRIPOD guidelines was 70.7%. Lack of external validation (14%) and clinical application (23%) of most included studies was detected. Reporting of model calibration was also missing in the majority of studies (5%).

Conclusion:

Implementation of various novel AI/ML tools promises significant advances in the classification, risk stratification, and management decisions for cancer pain. These advanced tools will integrate big health-related data for personalized pain management in cancer patients. Further research focusing on model calibration and rigorous external clinical validation in real healthcare settings is imperative for ensuring its practical and reliable application in clinical practice.

Keywords: Cancer pain; Cancer pain management; Machine learning; Artificial intelligence.

Introduction:

According to The International Association for the Study of Pain, pain is a subjective unpleasant sensation and distressing sensory or emotional experience related to tissue damage that can vary in intensity and duration [1, 2]. In cancer, pain represents a pervasive issue affecting a large proportion of cancer patients which significantly impacts their quality of life (QoL) and increases morbidity as well as mortality rates [2, 3]. More than 55% of cancer patients, especially those with advanced stages and metastasis, complain from pain [4, 5]. Additionally, 30% of cancer patients experience chronic pain, either because of the cancer itself or due to cancer treatment such as surgery, chemotherapy (CT), and radiation therapy (RT) [6]. The burden of pain in cancer patients is substantial, impacting their physical, psychological, and social well-being [4]. Moreover, untreated or poorly managed pain can lead to decreased QoL, functional impairment, and increased opioids prescription and healthcare costs [4-7]. Pain in people with cancer can manifest in various forms, including acute or chronic. Acute pain is usually temporary but severe while chronic pain can persist and last more than 3 months after starting point of pain throughout the cancer journey in cancer survivors, [8, 9].

The management of pain in cancer patients is a complex evolving field. It involves a multimodal approach that may include pharmacological interventions, such as opioids, non-opioid analgesics and adjuvant medications, as well as non-pharmacological approaches like physical therapy, psychotherapy, and interventional techniques [10]. Pain management in cancer patients follows the widely recognized and recommended World Health Organization (WHO) analgesics ladder, which consists of a three-step approach guiding healthcare providers in selecting and prescribing appropriate pain medications based on the severity of the pain, with the aim of providing effective pain relief while minimizing side effects. [11]. Overuse of unneeded high doses of opioids raises the risks of opioids' side effects and chronic abuse rates, which negatively affects patients' QoL [5]. Striking the right balance between pain relief and avoiding opioid-related adverse effects and potential abuse is a significant challenge in cancer pain management [4, 5, 10].

While healthcare providers follow the WHO pain management ladder, effectively controlling pain in cancer patients remains a significant challenge. There is an unmet need for accurate guidelines to help clinicians predict, classify, and manage pain in cancer patients. Novel tools are necessary to support clinicians in risk stratification and personalized pain management, particularly in the context of cancer-related pain.

In recent years, the healthcare industry has witnessed a surge in the adoption of artificial intelligence (AI) and machine learning (ML) tools. These technologies have demonstrated their potential to transform healthcare delivery, diagnosis, and treatment decision-making [12]. AI/ML tools can analyze vast amounts of data, identify patterns, and provide predictive insights, thereby aiding healthcare professionals in making informed decisions [12, 13]. ML is a subfield of AI that includes systems that 'learn from experience (E) with

respect to some class of tasks (T) and performance measure (P), if its performance at tasks T, as measured by P, improves with experience E' [14, 15].

Neural Networks are a subtype of ML algorithms, which can have different degrees of complexity and evolve into the so-called deep learning (DL) algorithms when they involve a large number of layers. Both ML and DL algorithms can be classed into supervised, unsupervised and reinforcement learning based on the way in which training data is presented to the model [14] (Figure 1). In supervised learning, the model is expected to learn the mapping between the training inputs and outputs presented during training to then be able to predict the outputs on an unseen input dataset. In unsupervised learning, the model is presented with unlabeled data and expected to learn the patterns and correlations by itself based on the input data only. In reinforcement learning, the algorithm learns to map inputs to actions by maximizing (or minimizing) a reward (or punishment) action evaluation signal. Additionally, AI algorithms can be grouped based on the task they are asked to perform: segmentation, regression or classification. Natural Language Processing (NLP) is an application of AI algorithms that is recently gaining interest in the clinical domain.

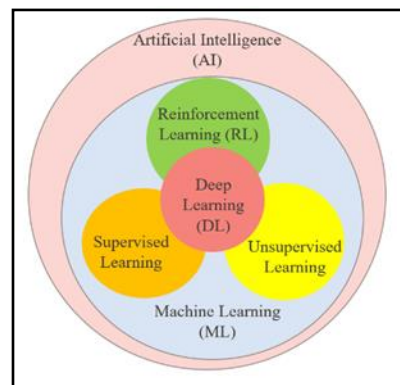


Figure 1: Types of Artificial Intelligence (AI) models.

AI/ML applications have begun to make inroads into the field of cancer pain prediction and management. These technologies offer the promise of more accurate pain assessment, personalized treatment recommendations, and improved patient outcomes [16]. However, the extent of their utilization and their impact on cancer pain prediction and management is an area that requires further investigation, and the results of the best performing models in cancer induced pain remain controversial and disperse.

In the dynamic landscape of AI/ML, the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) guidelines stand as a crucial framework for ensuring the transparency and reproducibility of predictive models. Developed to enhance the reporting quality of studies involving prediction models, TRIPOD provides a structured approach to the design, analysis, and interpretation of such models [17]. Adhering to these guidelines is paramount as it not only facilitates the effective communication of research findings but also fosters trust in the outcomes of predictive models. In the realm of ML, where complex algorithms increasingly influence decision-making across various health domains, the importance of adhering to TRIPOD guidelines is accentuated by the critical need for model validation [18]. Lastly, the validation process plays a pivotal role in assessing a model's generalizability and reliability, ensuring that its predictive capabilities extend beyond the training data [17, 18].

The performance and generalizability of an AI/ML prediction model is assessed by testing the model on a data subset that has not been used during the model training process.

The validation of a model can be internal or external, depending on whether the test dataset belongs to the same study cohort to that of the training dataset or is obtained from an entirely independent cohort. Both validation steps are crucial for the clinical implementation of a prediction model [18]. In particular, internal cross validation is a method to partition the data into training and testing subsets multiple times in order to provide a more accurate measure of model performance, especially with small datasets.

Comprehensive ML model performance assessment should include reporting of the model's discrimination ability and model calibration as part of the internal and external validation process [19, 20]. Model discrimination is a measure of the model's ability to differentiate between two classes (e.g., event vs. no event) given the available inputs [20]. Model calibration involves ensuring that the predicted probabilities align closely with the actual probabilities of events occurring [19].

Our scientific questions are:

- Which AI/ML algorithms are used in cancer pain research and cancer pain management?
- What are the applications of AI/ML models in cancer pain medicine?
- Which are the best performing models for cancer pain prediction and opioids optimization?
- To what degree do the existing AI/ML models in cancer pain prediction follow the TRIPOD guidelines with respect to model performance reporting?

This systematic review aims to answer these questions and address the current gap in knowledge by analyzing existing literature on the role of AI/ML in cancer pain prediction and management decision-making. We also provide an overview of common supervised and unsupervised learning techniques, their general characteristics, and specific applications in cancer pain research, either in cancer pain prediction, cancer treatment related pain or pain management and opioids decisions.

Materials and Methods:

Protocol Registration:

Registration of this systematic review in the international prospective register database of systematic reviews (PROSPERO), was done on 16 October 2023 [ID number: CRD42023469865] in the context of human health care.

Search Strategy and Study Eligibility:

We conducted a systematic search of Ovid MEDLINE, Ovid EMBASE, and Clarivate Analytics Web of Science, for publications in English from the inception of databases to September 7, 2023. The concepts searched included “Cancer”, “Pain”, “Pain Management”, “Pain Measurement”, “Analgesics”, “Opioids”, “Artificial Intelligence”,

“Machine Learning”, “Deep Learning”, “Expert System” and “Neural Networks”. Both subject headings and keywords were utilized. The terms were combined using AND/OR Boolean Operators. Animal studies, in vitro studies, and conference abstracts were excluded. The complete search strategy is detailed in Tables S1–S3.

Screening process:

Identified articles from the Search process were uploaded into the Covidence screening application [21], and screening through Covidence was conducted by two independent reviewers. Screening of the titles and abstract was first conducted and then screening of the full text was done on the retrieved articles. Final included articles were extracted from Covidence for full-text review and data collection.

Inclusion Criteria:

Studies eligible for inclusion in this systematic review had the following criteria 1) be published in English, 2) investigate AI/ML applications in cancer pain medicine or cancer pain management, and 3) involve human subjects.

Exclusion Criteria:

Articles were excluded if they met any of the following criteria: 1) out of scope of our study 2) no AI/ML application, 3) non-cancer pain study, 3) not an original study (i.e., review article, letter, conference abstract), 4) duplicate publication or correction of an original article, 5) descriptive study that didn't apply or test models.

Data Synthesis:

This study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines [22].

Data Collection Process:

The collected data included the following: AI/ML technique(s) used, input features and output variables, validation methods, and model performance metrics. Adherence to TRIPOD guidelines was analyzed using TRIPOD checklist [18]. Included articles were categorized according to the use of the AI/ML in cancer pain and grouped into the following groups: (1) models for prediction of cancer related pain intensity, cancer pain diagnosis or cancer pain initiation, (2) models for prediction of post cancer treatment pain (e.g., cancer surgery, RT or CT), (3) models for cancer pain management prediction or as a decision support system for cancer pain management.

Results:

Search and screening Results:

This comprehensive search resulted in identification of 436 studies through search of MEDLINE (n=189), EMBASE (n=189) and Web of Science (n=132). After screening and eligibility assessment, 283 articles were excluded. A total of 44 studies met the inclusion

criteria and were included in this review. The search strategy process is illustrated in a PRISMA flow diagram (Figure 2). All included studies were published between 2006 and 2023, with an increase in the number of publications for AI/ML studies in cancer pain research from 1 (2006-2009) to 26 (2020-2023) (Figure 3).

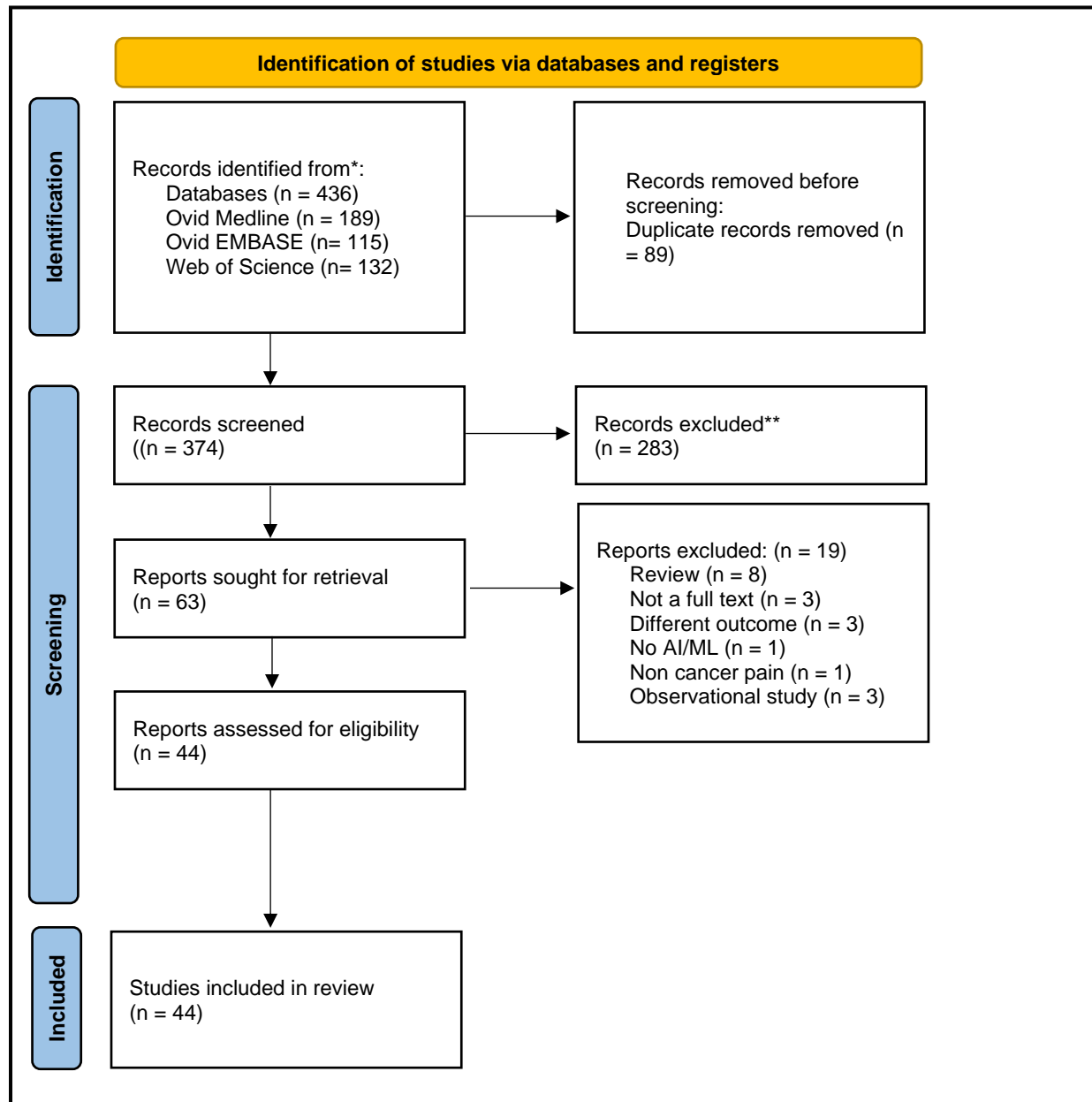


Figure 2: PRISMA Flow Diagram for systematic reviews of AI and ML in cancer pain research

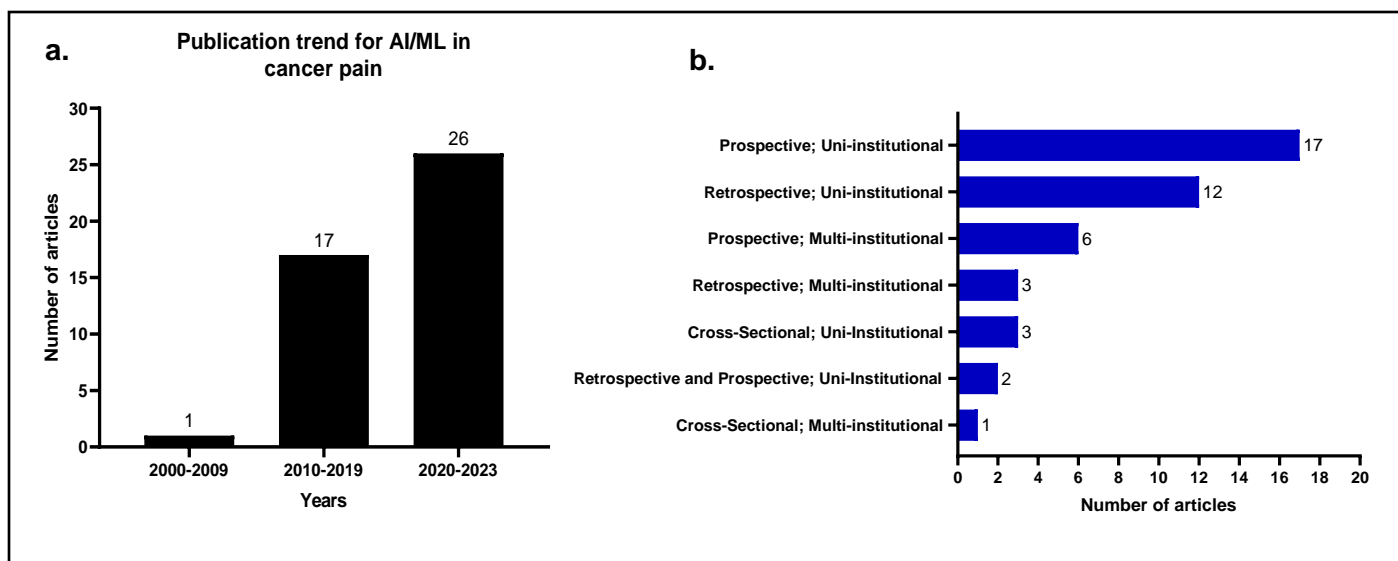


Figure 3: a. Publications trends for AI/ML models used for cancer pain research between 2006-2023. **b.** Types of studies and the number of articles per each type.

Design and populations:

Most studies used a prospective uni-institutional cohort to develop their models ($n=17$ studies) while the least used approach involved the cross-sectional multi-institutional cohort design ($n=1$) (Figure 3.b). The median sample size to build the models was 320 (range: 21-46104, IQR 140-1000) 95% CI 156-900. Five studies did not specify the size of the study population. In 95% ($n=23$) of studies the cohort size was between 100-1000 patients, while 23% ($n=9$) of studies used >1000 patients and 18% ($n=7$) used <100 patients.

AI/ML algorithms used in cancer pain research:

The most common AI/ML algorithms identified in this review, their general characteristics and specific use in cancer pain research are summarized in (Table 1). Some studies used a single model, while the majority of studies explored multiple models (each model was explored separately) (55%, $n=24$ articles) (Figure 4.a). The most common algorithms used in these multiple models studies were Random Forest (RF) and Logistic Regression (LR) ($n=15$ studies for each) (Figure 4.b). Other single model studies included Neural Networks (NN) (16% $n=7$), Decision Trees (DT) (7% $n=3$), Decision support system-Fuzzy ($n=2$), Clustering ($n=2$), LR ($n=2$), Bayesian networks ($n=1$), NLP ($n=1$) and 2 studies used novel models (Figure 4.a). Several other models used in multiple models studies (e.g., SVM, transformer, RF, GBM, Lasso, and NB) (Figure 4.b)

Table 1: AI/ML model characteristics and applications in cancer pain research.

| Machine Learning Algorithm | Characteristics | Cancer Pain Usage |
|--|---|--|
| Supervised Learning [23] | The output is labeled with the desired value (e.g., pain score or opioids dose) | |
| Classification | The output variable is a binary and/or categorical response | |
| Decision trees (DT) [24], [25], [26], [27], [28], [29], [16], [30], [31] | Generates understandable rules with both categorical and continuous variables for prediction and classification purposes. | <ul style="list-style-type: none"> • Cancer pain prediction • Post cancer treatment pain prediction • Cancer pain analgesics requirement |
| Random Forest (RF) [24], [32], [33], [34], [26], [35], [27], [36], [37], [16], [10], [38], [39], [40], [41] | An ensemble approach that combines the output of multiple decision trees to reach a single high accurate result. It provides a good predictive performance, low overfitting, and easy interpretability. | <ul style="list-style-type: none"> • Cancer pain prediction • Post cancer treatment pain prediction • Cancer pain analgesics requirement • Cancer pain related unscheduled healthcare prediction • Prediction of remote consultation/visits or patient satisfaction due to cancer pain. |
| Gradient boosting (GBM); extreme gradient boosting (XGBoost) [32], [33], [35], [27], [28], [37], [16], [38], [41] | A robust ensemble boosting algorithm that trains the model sequentially for prediction and classification purposes. It combines several weak learners into strong learners. | <ul style="list-style-type: none"> • Post cancer treatment pain prediction • Cancer pain related unscheduled healthcare prediction • Cancer pain analgesics requirement • Cancer persistent pain prediction |
| Naïve Bayes (NB) Classifier [26], [37], [16] | Used for binary or multi-class classification problems, based on Bayes rule and conditional independence assumption. | <ul style="list-style-type: none"> • Post cancer treatment pain prediction • Cancer persistent pain prediction |
| Regression | Predicts the probability using continuous data. | |

| | | |
|---|---|--|
| <p>Generalized linear mixed models (GLMMs), least absolute shrinkage and selection operator (Lasso), Linear regression (LR), Logistic regression (LR), Bayesian regression [42], [32], [43], [33], [34], [35], [27], [44], [45], [46], [39], [41], [31]</p> | <p>Model that estimates the relationship between one dependent variable and one or more independent variables using a line.</p> | <ul style="list-style-type: none"> • Cancer pain prediction • Post cancer treatment pain prediction • Cancer pain analgesics requirement • Cancer pain related unscheduled health care prediction • Association between epigenetics and persistent cancer pain |
| <p>Support vector machine (SVM) [43], [34], [47], [26], [27], [28], [37], [40], [41]</p> | <p>Type of supervised learning algorithm used to solve classification and regression tasks. Creates a hyperplane to separate two classes.</p> | <ul style="list-style-type: none"> • Cancer pain prediction • Post cancer treatment pain prediction • Cancer pain analgesics requirement |
| <p><u>Unsupervised Learning</u> [36]</p> | <p>Unlabeled and unclassified datasets used to train machines; Used to categorize unsorted data based on features, similarities and differences.</p> | <ul style="list-style-type: none"> • Mapping parameters and post cancer treatment persistent pain |
| <p>Clustering [25], [48], [49]</p> | <p>Machines divide the data into clusters based on features, similarities and differences.</p> | <ul style="list-style-type: none"> • Risk features identification for cancer pain • Identifying groups sharing patterns in post cancer treatment persistent pain |
| <p>Association [50]</p> | <p>Machines find interesting relations and connections among variables within large datasets that are input.</p> | <ul style="list-style-type: none"> • Correlation between genetic SNPs and post treatment cancer pain outcome. |
| <p><u>Neural Networks (NN)</u> [33], [25], [34], [26], [27], [28], [51], [52], [16], [10], [46], [38], [53], [54], [55], [56], [31]</p> | <p>Composed of node layers, containing an input layer, one or multiple hidden layers and an output layer. Could be supervised, unsupervised or semi-supervised.</p> | <ul style="list-style-type: none"> • Cancer pain prediction • Post cancer treatment pain prediction • Cancer pain identification using video analysis • Prediction of remote consultation/visits due to cancer pain. • Association between epigenetics and persistent cancer pain |

| | | |
|--|--|---|
| Deep Learning (DL) [57], [56] | Subclass of NN that Includes many layers of the neural network and massive volumes of complex and disparate data. | <ul style="list-style-type: none"> Genetic mutation associated cancer pain Cancer pain prediction |
| Natural Language Processing (NLP) [58], [59], [60] | A component of AI that uses the ability of a computer program to understand human language either written or spoken, it could be supervised or unsupervised. | <ul style="list-style-type: none"> Identification of cancer pain and cancer attributes |
| Decision Support System [61], [62], [31] | A subtype of AI using computerized programs to help decision making, judgment and actions in an organization. | <ul style="list-style-type: none"> Evaluate decision support computer program for cancer pain management. Pain treatment recommendations and staging Pain intensity identification |

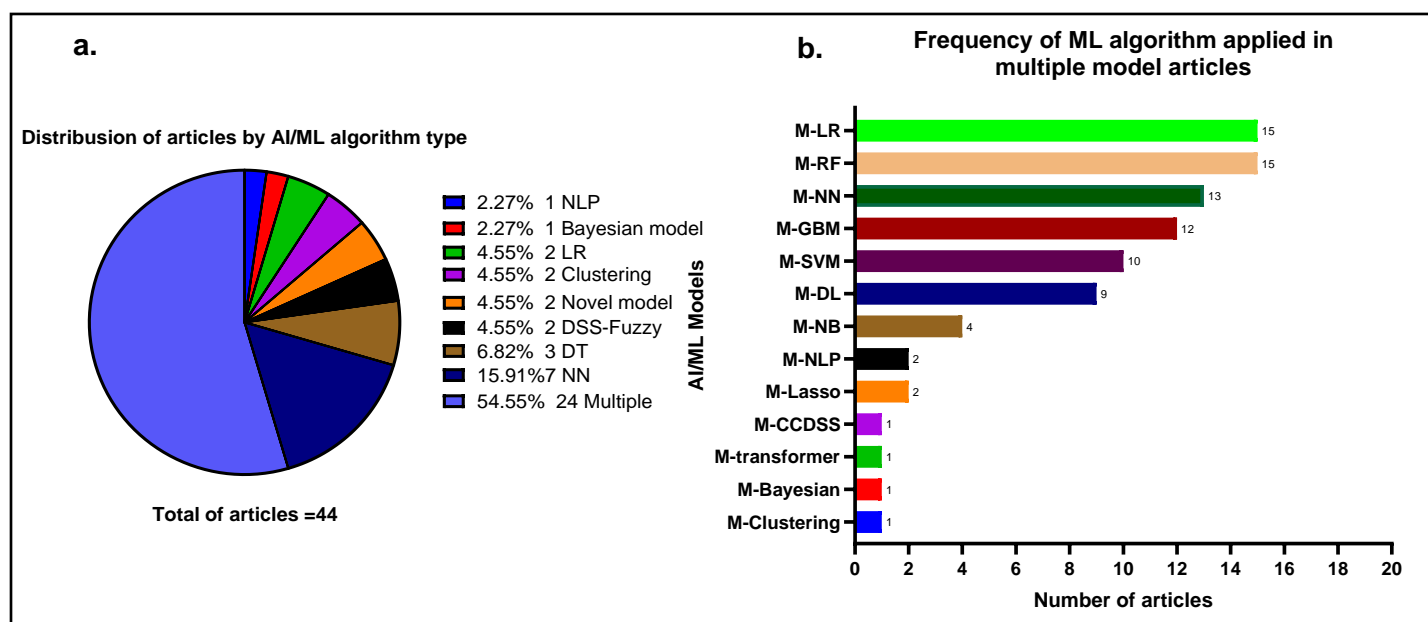


Figure 4: a. Distribution of articles by AI/ML algorithm type. **b.** 'Frequency of ML algorithm applied in multiple model articles' (*M*; *multiple*)

AI/ML Models used for cancer related pain research:

This review identified 18 studies describing the implementation of AI/ML in cancer pain research (Table 2).

In a study by Knudsen et al. (2011) [42] a cross-sectional analysis of 2278 cancer patients utilized Linear regression to identify key variables associated with pain, revealing breakthrough pain, psychological distress, sleep issues, and opioid dose as significant factors. Shimada et al. (2023) [29] employed DT in a retrospective study predicting end-of-life cancer patients' pain, showcasing the potential of ML in understanding complex patient scenarios. Cascella et al. (2023) [52] and Cascella et al. (2022) [10] explored video analysis and telemedicine, respectively, employing NN and various models to enhance cancer pain management. DiMartino et al. (2022) [58] used NLP to evaluate uncontrolled symptoms (i.e., pain), showing promise of AI in pain prediction (accuracy 61%). Lou et al. (2022) [39] delved into patient satisfaction in cancer pain consultation, using LR and RF, revealing associations with physician factors but facing limitations in data collection and ML metric reporting. Lu et al. (2021) [59] tested ML algorithms on young cancer survivors, with Bidirectional Encoder Representations from Transformers (BERT) showing higher accuracy in identifying pain interference. Moscato et al. (2022) [40] explored physiological signals for pain assessment. Results demonstrated the feasibility of using ML for pain assessment via physiological signals in real-world contexts. Heintzelman et al. (2013) [60] demonstrated the feasibility of text mining using NLP for predicting pain severity in metastatic prostate cancer. Miettinen et al. (2021) [30] identified pain phenotype clusters using DTs, while Xuyi et al. (2021) [55] predicted multiple symptoms using NNs. Akshayaa et al. (2019) [57] achieved high accuracy in detecting chronic pain with CNN, and Bang et al. (2023) [56] predicted breakthrough pain onset with DL models. Masukawa et al. (2022) [41] utilized LR, RF, GBM, and SVM to detect social distress and spiritual pain in palliative care. Pombo et al. (2016) developed a Computerized clinical decision support system (CCDSS) for pain intensity prediction, showing accuracy but limited clinical validation discussion. Xu et al. (2020) [63] explored herbal drugs' molecular mechanisms in pain subtypes. Pantano et al. (2020) [49] identified Break through cancer pain (BTcP) clusters.

These studies employ various AI/ML models, showcasing their potential in understanding patient satisfaction, identifying pain-related attributes, and predicting pain in cancer patients across different age groups and healthcare settings. Most studies used clinical and demographic variables as input; however, some studies applied models utilizing innovative inputs such as physiological signals [40], facial expressions [52], textual data [41], and even herbal categories [63], indicating the exploration of novel data sources. This diversity in inputs strengthens the versatility of the models. Some studies showcase innovation by assessing pain through video analysis of facial expressions [52], wearable devices capturing physiological signals [40], and even employing NLP for identifying pain severity in unstructured medical records [58]. These novel approaches strengthen the potential for non-traditional but effective pain assessment.

Most studies lacked detailed information on the extent of external validation of the AI models applied and the clinical application in real healthcare scenarios. Only 3 out of 18 studies performed external validation, and 7 studies discussed the clinical evaluation and application of the models used. Cascella et al, 2022 & 2023 [10] [38] applied external

validation and clinical evaluation of the models for the remote consultation prediction in cancer patients. Additionally, Cascella et al, 2023 [52] performed a clinical trial to test the NN using video recording of facial expression for discriminating between the absence and presence of pain in cancer patients. Heintzelman et al., 2013 [60] proved the feasibility and generalizability of NLP through an external validation using separate data source, furthermore, they evaluated the clinical application of the NLP using text mining for pain prediction in patients with metastatic prostate cancer. Pombo et al., 2016 [31] clinically evaluated the CCDSS compared with clinical advice for pain intensity, however, there was a lack of generalizability of the decision model due to the uni-institual cohort. Moscato et al., 2022 [40] applied a real-world context to develop an automatic pain assessment model, however, the small size cohort was a limitation with the need for bigger cohort for clinical validation.

Table 2: Characteristics of studies that used AI/ML algorithms in cancer pain research.

| Author, year | Population | Cohort study | AI/ML model (s) | Input variables | Output variable | Assessment tools | Aim/objective |
|---------------------------|--------------------------|--------------------------------------|-----------------|--|--|--|--|
| Knudsen et al., 2011 [42] | 2278 Cancer patients | Cross-Sectional; Multi-institutional | LR | 46 variables -Functional status -Cognitive function | 1. Average pain 2. Worse pain 3. Pain relief | 11-point Numerical Rating Scales (0-10) | Identify variables associated with pain |
| Shimada et al, 2023 [29] | 213 patients with cancer | Retrospective; Multi-Institutional | DT | -Patient clinical and demographic characteristics - performance status | Pain in cancer patients | Characteristic visual information | Using ML to predict pain in End of Life cancer patients |
| Cascella et al, 2023 [52] | Cancer patients | Prospective; Multi-Institutional | NN | -Facial expressions characteristic. - video recordings | Pain in cancer patients | video recordings. A set of 17 Action Units (AUs) was adopted. For each image | Constructed a ML model for discriminating between the absence and presence of pain in cancer patients using video analysis |

| | | | | | | | |
|-----------------------------|---|----------------------------------|--|--|---|---|---|
| Cascella et al., 2022 [10] | 158 cancer patients | Prospective; Uni-Institutional | RF GBM LASSO-RIDGE NN | -Demographic -Clinical variables -Therapeutics Drugs for pain, | Number of remote consultations due to pain | -Number of remote consultations -Pain severity assessment: Morphine dose | Development of ML predictive models for identifying patients who may require more remote consultations due to pain. |
| DiMartino et al., 2022 [58] | Cancer patients admitted to UNC Hospitals, a total of 1,644 hospitalizations were included. | Retrospective; Uni-institutional | NLP | -Clinical notes and records of symptom documentation -highest severity reported for pain | Pain severity labeled as: "controlled" (none, mild, not reported) or as "uncontrolled" (moderate or severe) | Common Terminology Criteria for Adverse Events (CTCAE) grade. | To evaluate NLP used to identify pain in EHR among hospitalized cancer patients. |
| Cascella et al., 2023 [38] | 226 cancer patients and 489 telemedicine visits for cancer pain management | Prospective; Multi-Institutional | DL (NN) RF GBM LASSO-RIDGE regression | -Demographic -Clinical variables -Background pain (nociceptive, neuropathic) and breakthrough cancer pain (BTcP) | The number of remote consultations" | The number of remote consultations recorded. | Investigate specific AI model efficacy for enhancing the telemedicine approach to cancer pain management. |
| Lou et al., 2022 [39] | Cancer patients with uncontrolled pain who participated in clinical consultations with physicians | Retrospective; Uni-institutional | LR RF | -Depth of prognosis discussion -Doctor's age, gender -Patient's race -Extent of shared decision making | Patient Satisfaction with provider visit/treatment | Questions asked by physician and standardized patients (SPs) | Use ML to examine the association between patient satisfaction and physician factors in clinical |

| | | | | | | | |
|-------------------------------|---|------------------------------------|--|--|--|--|---|
| | | | | | | | consultations about cancer pain. |
| Lu et al., 2021 [59] | Child and adolescent survivors of cancer, aged 8 to 17 years, | Cross-Sectional; Uni-Institutional | Transformers (BERT) NLP SVM XGboost | Meaning units in pain and fatigue as variables. | 1. Pain 2. Fatigue | Standard surveys with prespecified content of PROs | Test the validity of NLP and ML algorithms in identifying different attributes of pain experienced by child and adolescent survivors of cancer. |
| Moscato et al., 2022 [40] | 21 cancer patients | Prospective; Uni-institutional | SVM RF MP LR AdaBoost | Physiological signals (e.g., photoplethysmography, electrodermal activity) | Pain Experience (Pain ratings) | Edmonton Symptom Assessment Scales (ESAS) for pain assessment. | Develop an automatic pain assessment method based on physiological signals recorded by wearable devices. |
| Heintzelman et al., 2013 [60] | 33 men with metastatic prostate cancer | Retrospective; Uni-institutional | NLP LR | -Electronic, radiologic, RT, and pathology medical records Pain severity -Factors associated with pain experience (e.g., receipt of opioids and palliative radiation -Patient demographics (age, race/ethnicity). | Pain Phenotype intensity: 1. Severe Pain 2. No Severe or controlled Pain | Pain categorization model based on a conservative four-tiered pain scale: no pain (category 0); some pain (category 1); controlled | To test the feasibility of using text mining to predict pain in patients with metastatic prostate cancer. |

| | | | | | | | |
|-----------------------------|--|----------------------------------|---|--|---|--|--|
| | | | | | | pain (category 2); severe pain (category 3). | |
| Miettinen et al., 2021 [30] | 320 cancer patients with persistent pain | Prospective; Multi-institutional | DT (CART and PART) | -Pain Phenotype- Intensity related features -Pain etiology-related information -Psychological parameters -Demographic parameters -Lifestyle-related parameters, - Previous treatments -Comorbidities. | 1. Pain Intensity 2. Number of Pain areas 3. Pain Duration 4. Activity Pain Interference 5. Affective Pain interference | The Brief Pain Inventory (BPI) for assessing pain intensity. | (1) identify patterns arising from different pain phenotypic parameters (2) selecting parameters in associating a patient with a particular pain phenotype. |
| Xuyi et al., 2021 [55] | Cancer patients between 2008 and 2015. [training and test cohorts consisted of 35,606 and 10,498 patients, respectively] | Retrospective; Uni-institutional | NN | 39 unique covariates: -Demographics, -Clinical features, -Treatment characteristics -Baseline PRO -Health care utilization measures. | 1. Severe Pain 2. Moderate to severe depression 3. Poor Well-being | ESAS; and interRAI | Predict the risk of Severe Pain in cancer patients |
| Akshaya et al., 2019 [57] | 900 Images of Cancer patients experiencing chronic pain and with ZFH2 mutations. | Prospective; Uni-Institutional | Deep Convolutional Neural Network (CNN) | -Pathological images with ZFH2 mutation -epigenomic alterations (DNA methylation), | Chronic pain | -Not specified | Designing an early mutation detection tool to identify the presence of pain inducing gene ZFH2 |

| | | | | | | | |
|----------------------------|---------------------------------|------------------------------------|---|--|---|--|--|
| | | | | | | | using deep CNN. |
| Bang et al., 2023 [56] | 34,304 cancer patients. | Retrospective; Uninstitutional | Transformer CNN (DL) | -Pain intensity scores - lengths and time | Cancer Pain Exacerbation | 11-point Numerical Rating Scales (0-10) | Investigate the clinical relevance of DL models that predict the time of breakthrough pain onset in cancer patients. |
| Masukawa et al., 2022 [41] | 808 patients who died of cancer | Retrospective; Uninstitutional | LR RF GBM SVM | -Unstructured text data contained in EMRs. -Social distress -Spiritual pain Severe physical/psychological symptoms. | 1. Social Distress 2. Spiritual Pain 3. Severe Physical and Psychological Symptoms (e.g., pain) | Japanese version of the Support Team Assessment Schedule (STAS-J). | Develop models to detect social distress, spiritual pain in terminally ill cancer patients. |
| Pombo et al., 2016 [31] | 32 cancer volunteers | Prospective; Uninstitutional | CCDSS ANN Bayesian algorithms LR DT | Clinical data, Treatment protocol Patient reported data | Mean pain intensity | -Not specified | To develop and validate a CCDSS as a method used pain evaluation system for pain intensity prediction. |
| Xu et al., 2020 [63] | Not specified | Retrospective; Multi-institutional | Herb Networks (SHN) and (THN) | -Pain-related herbs -Herbal molecules -Human and gut microorganism targets -Pathways | 1. Cancer pain 2. Chronic cough related neuropathic pain 3. Reproduc | Not specified | Understanding of the molecular mechanisms of pain subtypes that |

| | | | | | | | |
|---------------------------|--|--------------------------------|------------------|---|----------------------------------|---|--|
| | | | | Herb categories (HC1, HC2, HC3) | tion and autoimmune related pain | | herbal drugs are participating |
| Pantano et al., 2020 [49] | 4,016 cancer patients suffering from BTcP. | Retrospective; Uninstitutional | Cluster analysis | Eight BTcP-defining variables (e.g., number of episodes, peaks duration, type, intensity, etc.) | Break through cancer pain (BTcP) | 11-point Numerical Rating Scales (NRS) (0-10) | Explore distinct subtypes of BTcP using an unsupervised learning algorithm |

AI/ML Models used for cancer treatment induced pain research:

Nineteen articles focused on implementing AI/ML algorithms in cancer treatment pain research were detected in our review. The characteristics of these studies are summarized in Table 3.

The study by Chao et al. in 2018 [24] aimed to predict radiation-induced chest wall pain in Non-Small Cell Lung Cancer (NSCLC) patients using DT and RF models. The study demonstrated that ML models are predictive for chest wall pain toxicity after RT in Lung cancer patients. Additionally, Olling et al. (2018) [43] conducted a study aimed at predicting pain while swallowing (odynophagia) post-RT in lung cancer, applying LR, SVM and Generalized Linear Models (glmnet) based on various parameters. These models accurately predicted odynophagia during lung cancer RT, revealing their effectiveness in pain prediction during the course of treatment for lung cancers. Studies conducted by Sun et al. (2023) [32], Juwara et al. (2020) [33], Lotsch et al. (2017) [23], Lotsch et al. (2018) [36], Sipila et al. (2012) [44], Lotsch et al. (2018) [34] and Wang et al. (2021) [27], investigated various ML models for persistent/chronic pain prediction and features identification in post-surgery in breast cancer patients. Studies by Sun et al. (2023) [32] and Juwara et al. (2020) [33] revealed that the novel ML approaches, including RF, GBM, and XGBoost, exhibited higher performance in predicting Chronic Postsurgical Pain (CPSP) over the traditional regression models. Lotsch et al. (2017) [23] revealed that supervised classification ML techniques robustly excluded post-surgical persistent pain, exhibiting a high accuracy of 94.4% based on preoperative cold pain sensitivity. Sipila et al. (2012) [44] developed a screening tool using ML to identify risk factors contributing to persistent post-surgery pain using a Bayesian model. The study's ML model showed high predictive performance and identified significant predictors of prolonged pain after breast cancer surgery, such as preoperative chronic pain and multiple previous operations. Lotsch et al. (2020) [48] used automated evolutionary algorithms, hierarchical clustering, and LR to identify distinct pain patterns from 6 to 36 months post-surgery and revealed

three unique patient groups exhibiting differing persistent postoperative pain patterns. Wang et al. (2021) study results demonstrated that XGBoost algorithm exhibited the highest precision in predicting chronic pain, while the Naive Bayes (NB) model showed commendable performance in recall rate and F1 score. Sipila et al. (2020) [25] used various ML techniques to identify psychological factors influencing persistent post-treatment pain. They found that psychological and sleep-related parameters played a crucial role in grouping patients dealing with persistent pain after breast cancer treatments.

In a retrospective and prospective study, Guan et al. (2023) [26] investigated postoperative pain in hepatocellular carcinoma (HCC) patients after trans arterial chemoembolization (TACE). Using various ML models, they found that the RF model accurately predicted the risk of pain following TACE in HCC patients. Barber et al. (2022) [35] investigated a group of 34 women after gynecologic cancer surgery, utilizing ML models such as LR, RF, GBM, and XGBoost. They employed patient-reported outcomes (PROs) and wearable device data to predict the specific day of unscheduled healthcare utilization events, with a particular emphasis on pain intensity. The study's primary finding was the feasibility of a PRO-based program in accurately predicting the occurrence of unscheduled healthcare utilization following surgery, especially concerning pain.

Reinbolt et al. (2018) [50] examined a novel analytic algorithm (NAA) to analyze genetic single nucleotide polymorphisms (SNPs) for predicting Aromatase Inhibitor Arthralgia (AIA) in breast cancer patients treated with Aromatase Inhibitors (AIs). Their key finding was the identification of 70 SNPs from 57 genes that predicted AIA with a notable accuracy of 75.93%. The study demonstrated the potential of using genetic markers to anticipate AIA, providing a promising avenue for preemptive risk identification. Furthermore, Kringel et al. (2019) [46] utilized kNN, SVM, LR, and NB, their ML analysis to understand the association between DNA methylation of glial/opioid-related genes and persistent post-surgery pain in breast cancer patients. Their study highlighted that global DNA methylation held a comparable diagnostic accuracy for persistent pain, compared to established non-genetic predictors. Finally, Lotsch et al. (2022) [53] explore the potential of ML techniques such as (PCA, NN, RF) to identify proteins that distinguish patients with persistent postsurgical neuropathic pain (PPSNP). The study revealed 19 proteins that showed significant differences between the groups, marking a key finding.

As for the previous set of studies, most studies considered in this subgroup also lacked detailed information on the extent of external and clinical validation in real healthcare scenarios. Three out of 19 studies, used external clinical validation of the AI/ML models. Guan et al., 2023 and Wang et al., 2021 [26] [28] used an external prospective clinical cohort for clinical evaluation of the models. Additionally, Im et al., 2006 [61] used external cohorts of nurses from the internet for the clinical evaluation of the decision support computer program (DSCP) developed for pain prediction and management. Although studies used genomic, epigenomic and proteomic data to build the classification and clustering pain models, got the identified molecules from the clinical settings, were

previously identified, however, no actual clinical validation and application were used, and the clinical evaluations of the models are still needed.

Table 3: Characteristics of studies used AI/ML algorithms in cancer treatment related pain research.

| Author, year | Population | Cohort study | AI/ML model (s) | Input variables | Output variable | Assessment tool | Aim/objective |
|--------------------------|--|----------------------------------|---|--|--|--|--|
| Chao et al., 2018 [24] | 197 cancer patients with Stage I NSCLC treated with SBRT | Prospective; Uninstitutional | DT RF | 25 patient, tumor and dosiomic features | Radiation induced chest wall pain, Chest wall syndrome (CWS) | CTCAE v4 grade ≥ 2 chest wall pain | Utilize ML algorithms to identify chest wall pain RT toxicity predictors to develop dose-volume constraints |
| Sun et al., 2023 [32] | 1152 patients with primary breast cancer undergoing mastectomy | Prospective; Multi-Institutional | RF GBM LR XGBoost | 6 leading predictors including (pain score, post-menopausal status, urban medical insurance, history of at least one operation, under fentanyl with sevoflurane general anesthesia, and received axillary lymph node dissection) | Chronic postsurgical pain (CPSP) at 12 months after surgery | - Modified Brief Pain Inventory > 0 -2016 International Association for the Study of Pain (IASP) criteria | Develop prediction models for CPSP after breast cancer surgery using ML approaches |
| Olling et al., 2018 [43] | 131 NSCLC cases received RT | Retrospective; Uninstitutional | LR SVM Generalized Linear Models (glmnet) | Clinical, tumor and dose volume parameters. | Pain while swallowing (odynophagia) after RT in lung cancer | Semi-structured weekly in-person interviews reported by RT | Generate predictive ML models for odynophagia needing prescription pain medication during lung RT lung cancer using nursing workflow |

| | | | | | | | |
|--------------------------|--|--|--|---|--|--|--|
| | | | | | | Technologist Nurse (RTN) | |
| Juwara et al., 2020 [33] | 195 female patients scheduled to undergo breast cancer surgery | Prospective; Uninstitutional | Solitary Least square regression (OLS) Ridge regression (RR) Elastic net (EN) RF GBM NN LR | Demographic and Clinical features | Neuropathic pain after breast cancer surgery | 4 interview scores (DN4-interview; range: 0–7)- 3 months after surgery | Assess the utility of ML models developed as a tool to predict pain and identify specific pain characteristics after breast cancer surgery |
| Sipila et al., 2020 [25] | 337 women treated for breast cancer | Cross-Sectional; Uninstitutional | Cluster analysis NN DT | - Psychological and sleep-related parameters - Parameters related to pain intensity and interference | Persistent pain in breast cancer survivors | Brief Pain Inventory (BPI) for pain. 11-point Numerical Rating Scales (NRS) (0-10) | Identify psychological features that may influence poorer coping with persistent pain |
| Guan et al., 2023 [26] | Retrospective 857 patients and prospective 368 patients with HCC received TACE | Retrospective and Prospective; Uninstitutional | RF SVM NN Naïve Bayes DT | 24 candidate variables, Demographics, clinical, tumor, and surgery data | Postoperative pain in HCC | NR: 0 points, no pain; 1-3 points, mild pain; 4-6, moderate pain; 7-10, severe pain. | To develop an early ML model for predicting pain after TACE in patients with HCC |

| | | | | | | | |
|----------------------------|---|---------------------------------|--|---|---|--|--|
| Lotsch et al., 2017 [23] | 900 women who were treated for breast cancer | Retrospective; Un-Institutional | Supervised classification ML techniques | Parameters acquired during the cold pain tolerance test | Presence or absence of persistent pain after breast cancer surgery, in 12-36 months | -11-point Numerical Rating Scales (NRS) (0-10) -Tonic cold pain test | Use Supervised ML techniques to test how accurately the patients' performance in a preoperatively performed tonic cold pain test could predict persistent post-surgery pain in breast cancer |
| Barber et al., 2022 [35] | 34 women after gynecologic cancer surgery | Prospective; Un-institutional | LR RF GBM XGBoost | -Clinical features -PRO responses -wearable device output | The day of an unscheduled health care utilization event using PRO parameters (e.g., Pain intensity) | Patient - Reported Outcomes Measurement Information System (PROMIS) | Test the feasibility of implementing a postoperative monitoring program for women with gynecologic cancers composed of PROs and a wearable activity monitor. And predict the day of unscheduled health care utilization for pain |
| Reinbolt et al., 2018 [50] | Patients with Breast Cancer received Aromatase inhibitors | Retrospective; Un-Institutional | ML: novel analytic algorithm (NAA)-Correlation network | -Genetic SNPs - Demographics features. | Aromatase inhibitor arthralgia (AIA) in breast cancer patients | AIA-positive and -negative patients, we ranked based on their discriminatory power. | Evaluate the potential of a NAA to predict AIA using germline SNPs data obtained before treatment initiation |
| Im et al., 2006 [61] | 122 nurses working with cancer patients | Retrospective; Un-Institutional | Decision support computer program (DSCP) | Age, Ethnic and socioeconomic characteristics variables | Usage profile, accuracy, and acceptance of the DSCP | - Accuracy: measured by determining whether the decision support was appropriate and | To evaluate a DSCP for cancer pain management |

| | | | | | | | |
|--------------------------|---|---|---|---|---|---|---|
| | | | | | | accurate - Satisfaction. (satisfaction score on a 1-10 scale) | |
| Wang et al., 2021 [28] | 746 lung cancer patients with bone metastasis | Retrospective; and prospective; Uni-Institutional | DT XGBoots SVM Bayesian NN (BNN) | - Demographic and Clinical data. -The driver gene of lung cancer -Five differentially expressed proteins of bone metastases -VAS pain assessment | Local treatment in lung cancer pain with bone metastases to decrease pain. output | -Visual analog scale (VAS) - Quality of life (QoL) scores Bone Metastases Module (EORTC QLQ-BM22) | Developed and validate ML models to predict patients who should receive local treatment to reduce pain in lung cancer patients with bone metastasis |
| Lotsch et al., 2018 [36] | 1000 breast cancer patients undergoing breast surgery | Prospective; Uni-Institutional | Both supervised and unsupervised mapping ML models 1.A supervised classification 2.A symbolic rule-based classifier | 542 different variables (clinical, demographics, medical history, medication, pre-op pain, opioids, tumor data) | Persistent pain after surgery in breast cancer patients | 11-point Numerical Rating Scales (NRS) (0-10) | Identify parameters that predict persistence of significant pain in breast cancer after surgery using ML models |
| Sipila et al., | 489 breast cancer treated with surgery | Prospective; Uni- | Bayesian model | - Demographics | Persistent pain after | 11-point Numerical | Develop a ML screening tool to identify presurgical |

| | | | | | | | |
|--------------------------|--|--------------------------------|---|---|--|---|--|
| 2012 [44] | | Institutional | | -Clinical -Vitals - treatment/ Hormonal , Chemo, radio therapy -Type of surgery -Pre- surgery pain - Psychological features | surgery in 6 months | cal Rating Scales (NRS) (0-10) | factors that predict persistence of pain after 6 months from surgery in breast cancer |
| Lotsch et al, 2020 [48] | 763 women treated with surgery for breast cancer | Prospective; Uni-Institutional | Unsupervised automated evolutionary (genetic) algorithms Hierarchical clustering LR | Patients' temporal features of NRS ratings | Clusters or groups of patients sharing patterns in the time courses of pain between 6 and 36 months after breast cancer surgery. | 11-point Numerical Rating Scales (NRS) (0-10) | Identify subgroups of patients sharing similar time courses of postoperative persistent pain in breast cancer, using ML algorithms |
| Lotsch et al, 2018 [64] | 1000 Breast cancer received surgery | Prospective; Uni-Institutional | Supervised ML implemented as RF | Psychological factors (depressive symptoms, state and trait anxiety, and anger inhibition) | Persistent pain after breast cancer surgery (PPSP) | 11-point Numerical Rating Scales (NRS) (0-10) | Create a simple questionnaire with a good ML predictive power for persisting pain after surgery in breast cancer patients. |
| Kringel et al, 2019 [37] | 140 women undergoing breast cancer surgery | Prospective; Uni-institutional | RF Adaptive boosting kNN | Next-generation sequencing for selected 77 genes. | Persistent Pain vs Non-persistent pain after breast cancer surgery | 11-point Numerical Rating Scales (NRS) (0-10) | Assessment whether NGS-derived genotypes were associated with persistent pain in patients who were treated with breast cancer surgery. |

| | | | | | | | |
|---------------------------|---|----------------------------------|--|--|--|---|--|
| | | | Naïve Bayes SVM LR | | | | |
| Wang et al., 2021 [16] | 3489 Breast Cancer patients | Prospective; Un-Institutional | DT RF XGBoost Naive Bayes (NB) CNN | Age, endocrine therapy, radiotherapy, and chemotherapy | Pain in post-operative breast cancer patients | -Not reported | Investigate the factors influencing chronic pain after radical mastectomy in breast cancer patients, using AI/ML algorithms |
| Kringel et al., 2019 [46] | 140 women who had undergone breast cancer surgery | Prospective; Multi-Institutional | kNN SVM LR Naïve Bayes | Global DNA methylation. DNA methylation status of TLR4, OPRM1, and LINE1 | Persistence of postoperative pain. | 11-point Numerical Rating Scales (NRS) (0-10) | Examine the association between the DNA methylation of two key genes of glial/opioid intersection and persistent postoperative pain. |
| Lotsch et al., 2022 [53] | 57 women who had undergone breast cancer surgery | Prospective; Un-Institutional | Supervised and Unsupervised PCA NN RF | Proteomics: patterns in 74 serum proteomic marker | Persistent postsurgical neuropathic pain (PPSNP) | -11-point Numerical Rating Scales (NRS) (0-10), sensory examination to diagnose PPSNP. -Brief Pain Inventory (BPI) | Identify proteins that are most informative in identifying patients with and without PPSNP after breast cancer surgery |

AI/ML Models used for cancer pain management prediction and decision:

Seven articles focused on implementing AI/ML algorithms in cancer pain management research were detected in our review. The characteristics of these studies are summarized in Table 4.

Kumar et al. (2023) [34] to create and validate AI/ML predictive models for postoperative fentanyl analgesic demand and associated outcomes. Employing generalized linear regression models (GLM), Linear-SVM, RF, and Bayesian regularized neural network (BRNN), their research amalgamated nongenetic clinical and genetic factors (SNPs), cold pain test (CPT) scores, and pupillary response to fentanyl (PRF) to predict 24-hour postoperative fentanyl needs, pain scores, and time for the first analgesic. The models demonstrated varied R-squared values (0.313 for SVM—Linear, 0.434 for SVM—Linear, and 0.532 for RF), indicating moderately good predictive capacity for these outcomes. Olesen et al. (2018) [47] conducted a study to predict oral morphine equivalent dose (MED) in cancer patients using SVM based on 18 SNPs within genes linked to opioid receptors and metabolic pathways. The SVM analysis found no significant associations between the specific genetic variants and the required opioid dose in cancer pain patients. While Bobrova et al. (2020) [27] conducted a study involving 90 pancreatic cancer patients utilizing fentanyl transdermal therapeutic system (TTS) for pain relief. The research employed diverse ML models and examined 57 genetic and non-genetic factors to predict pharmaco-resistance of fentanyl. The study identified the SVM as the most effective model for classifying pharmaco-resistance. Facciorusso et al. (2019) [51] conducted a study with 156 pancreatic cancer patients undergoing repeat celiac plexus neurolysis (rCPN) using ANN and LR to predict pain response post-rCPN. The study found that the ANN outperformed the LR model in predicting treatment response. Dolendo et al. (2022) [45] analyzed 148 cancer patients who underwent mastectomy, utilizing LR to assess factors affecting postoperative opioid use. The study suggested that these models, if implemented, could significantly impact preoperative counseling and patient satisfaction by aiding in the prediction of postoperative pain. Im et al. (2011) [62] involved 428 ethnic minority cancer patients to evaluate a Decision Support Computer Program for Cancer Pain Management (DSCP-CA). The DSCP-CA was developed to assist nurses in managing cancer pain among ethnic minority patients, potentially enhancing care in this specific patient population. Sokouti et al. (2014) [54] created an ANN model for pain management using thermal and electrical stimulations and time parameters to address total pain intensity and pain quality. The main finding revealed the ANN's accurate emulation of clinical data derived from stimulations, effectively managing severe pain in patients. A lack of external validation and clinical evaluation of the models was also detected in all studies involved in this subgroup.

Table 4: Characteristics of studies used AI/ML algorithms in cancer pain management prediction and decisions research.

| Author, year | Population | Cohort study | AI/ML model (s) | Input variables | Output variable(s) | Assessment tool(s) | Aims/Objectives |
|---------------------------|---|------------------------------------|---|---|--|---|---|
| Kumar et al., 2023 [34] | 257 cancer patients undergoing major breast surgery. | Cross-Sectional; Uni-institutional | Generalized linear regression model (GLM) SVM—Linear RF Bayesian regularized neural network (BRNN) | -Nongenetic clinical and genetic factors (SNPs) -Cold pain test (CPT) scores -Pupillary response to fentanyl (PRF) | 1. 24-hour fentanyl requirement 2. 24-hour pain scores 3. Time for first analgesic (TFA) in the postoperative period | -Cold pain testing -11-point Numerical Rating Scales (NRS) (0-10) -Pupillary response to fentanyl (PRF) using a linear Ultrasound probe (Sonosite, WA). | Develop and validate robust AI/ML predictive models for postoperative fentanyl analgesic requirement and other related outcomes |
| Olesen et al., 2018 [47] | 1237 cancer pain patients | Retrospective; Multi-institutional | SVM | 18 SNPs within the μ and δ opioid receptor genes and the catechol-O-methyltransferase gene | Required opioid dose in cancer pain patients. in oral MED | Oral morphine equivalent dose (MED) | Predict required opioid dose in cancer pain patients, using genetic profiling |
| Bobrova et al., 2020 [27] | 90 pancreatic cancer patients received fentanyl TTS for pain relief | Prospective; Uni-institutional | LR k nearest neighbors' algorithm (KNC) RF GBM DT NN SVM | 57 genetic and non-genetic factors: (13 genetics, 20 clinical, demographic, type of surgical treatment, 24 laboratories) | Development of Pharmacoresistance of Fentanyl in pancreatic cancer patients with chronic pain | Pharmacoresistance: according to the Naranjo scale Quality of life was assessed using the Palliative Medicine Symptom Rating Scale (ESAS), and cognitive | Develop a calculator for personalized risk assessment of opioid-associated drug resistance in patients with pancreatic cancer using fentanyl transder |

| | | | | | | | |
|-------------------------------|---|------------------------------------|---|---|--|---|---|
| | | | | | | functions were assessed using the Mental Status Assessment Scale (MMSE). | mal therapeutic system (TTS) as an example. |
| Facciorusso et al., 2019 [51] | 156 pancreatic cancer patients treated with repeat celiac plexus neurolysis | Retrospective; Uninstitutional | Artificial NN (ANN) LR | Baseline demographics, clinical and treatment rCPN characteristics | Pain response after repeated celiac plexus neurolysis (rCPN) | Visual Analogue Scale (VAS), ranging from 0 (no pain) to 10 (maximal pain) | Build an artificial NN model to predict pain response in pancreatic cancer patients received rCPN |
| Dolendo et al., 2022 [45] | 148 cancer patients that underwent mastectomy | Retrospective; Multi-institutional | LR | Patient demographics, clinical and surgical characteristics | Total opioid use on postoperative day 1 | Oxycodone milligram equivalents (OME) | Identify risk factors and develop ML-based models to predict patients who are at higher risk for postoperative opioid use after mastectomy. |
| Im et al., 2011 [62] | 428 cancer patients (ethnic minority cancer patients) | Prospective; Uninstitutional | Decision support computer system (DSCP) | - Demographics and socioeconomics (e.g., Ethnicity/race Sex) -Cancer pain experience | Pain Treatment decisions | -5 cancer pain scales (the visual analog scale [VAS], the verbal descriptive scale [VDS], the Wong-Baker Faces Pain Scale [FS], the Brief Pain Index- | Develop a DSC to support nurses' decisions about cancer pain management |

| | | | | | | | |
|---------------------------|--------------------|--------------------------------|-----|---|--|--|---|
| | | | | | | Short Form [BPI-SF], and the McGill Pain Questionnaire-Short Form [MPQ-SF]) -The Memorial Symptom Assessment Scale [MSAS], and the Functional Assessment of Cancer Therapy Scale [FACT-G]. | |
| Sokouti et al., 2014 [54] | 70 cancer patients | Prospective; Uni-Institutional | ANN | Thermal and electrical stimulations, and the time parameter | Total pain intensity management (Pain quality) | 11-point Numerical Rating Scales (NRS) (0-10) | Model an AI system to be more biologically efficient in pain quality modulation |

Performance of cancer related pain and pain management AI/ML models:

The main discrimination performance evaluation metrics of the AI/ML models were accuracy, receiver operating curve-area under the curve (ROC-AUC), sensitivity, specificity and root mean square errors. Several studies did not report any outcome performance of the models used. Table 5 summarizes the discrimination performance results and includes an assessment on the validation methods used and whether model calibration was assessed for the studies that applied AI/ML models in cancer pain studies.

Median AUC across studies reported the models AUC performance was 0.77% (range 0.47-0.99). Models used for cancer related pain studies showed the highest AUC with median AUC 0.86 (range 0.50-0.99), while the median AUC across studies of AI/ML models for cancer treatment related pain research and cancer pain management was 0.71 (range 0.47-0.89) and 0.80 (range 0.76-0.94), respectively (Figure 5.a). The RF model showed the highest median AUC (0.81, range 0.58-0.99), while the Lasso model showed the lowest median AUC (0.70, range 0.50-0.79). Median AUC for SVM (0.808,

range 0.71-0.87), NB (0.80, range 0.67-0.803), LR (0.78, range 0.65-0.86), NN (0.79, range 0.65-0.98), DT (0.76, range 0.58-0.89) and boosting (GBM) (0.71, range 0.56-0.89) (Figure 5.b). The Lasso model demonstrated the highest sensitivity (1.00), while the lowest specificity (0.00). SVM demonstrated the highest median specificity (0.74, range 0.52-0.97).

Only 2 studies (Sun et al., 2023 [32] and Juwara et al., 2020 [33]), out of all articles reported the calibration of the AI/ML models (Table 5). Sun et al., used the integrated calibration index, E50, E90 and Hosmer-Lemeshow (H-L) test for calibration of the models used in the study, where XGBoost model showed better performance than multivariable LR model with ICI, 0.05 (95% CI (0.038-0.122)) which was lower than LR ICI, 0.07 (95% CI (0.050 – 0.146)) and the other models; RF and GBM didn't show superiority in ICI and H-L P values reported for all models were (LR: 0.059, RF:0.429, GBM: 0.384 and XGBoost: 0.829) [32]. Jawara et al., generated the calibration plot of the predicted probability against the observed probabilities. The MCA adjusted model showed good calibration. The MCA unadjusted logistic classifier, the intercept: 0.02 (-0.50,0.47) and slope 0.98 (0.42,1.58), while for the MCA adjusted logistic classifier, the intercept: 0.03 (-3.4,0.34) and slope: 1.00 (0.40, 1.60) [33].

Table 5: Models performance, validation and calibration reported in included studies.

| Study | AI/ML model | Accuracy | AUC | Sensitivity/ Recall | Specificity/ Precision | Root mean square error | Validation | Calibration |
|---|-------------|----------|-------|---------------------|------------------------|------------------------|--------------------------|--------------|
| AI/ML Models used for cancer related pain research | | | | | | | | |
| Shimada et al, 2023 [29] | DT | 0.685 | 0.582 | 0.849 | 0.241 | Not reported | 10-fold Cross-Validation | Not Reported |
| Cascella et al, 2023 [52] | NN | 0.9448 | 0.98 | 0.968 | 0.9528 | Not reported | Train/test | Not Reported |
| Cascella et al., 2022 [10] | RF | 0.7 | 0.98 | 0.69 | 0.71 | Not reported | 8-fold cross-validation | Not Reported |
| | GBM | 0.5 | 0.59 | 0.69 | 0.29 | Not reported | | |
| | LASSO-RIDGE | 0.53 | 0.5 | 1 | 0 | Not reported | | |
| | ANN | 0.57 | 0.95 | 0.5 | 0.64 | Not reported | | |

| | | | | | | | | |
|------------------------------|----------|--------------|--------------|--------------|--------------|--------------|----------------------------|--------------|
| DiMartino et al., 2022 [58] | NLP | 0.61 | Not reported | 0.69 | 0.46 | Not reported | 10-fold cross-validation | Not Reported |
| Cascella et al., 2023 [38] | RF | 0.8 | 0.99 | 0.69 | 0.71 | Not reported | K-fold cross-validation | Not Reported |
| | GBM | 0.62 | 0.87 | 0.69 | 0.29 | Not reported | | |
| | LASSO | 0.57 | 0.7 | 1 | 0 | Not reported | | |
| | ANN | 0.71 | 0.92 | 0.5 | 0.64 | Not reported | | |
| Lou et al., 2022 [39] | RF | Not reported | 0.96 | Not reported | Not reported | Not reported | Not reported | Not Reported |
| | LR | Not reported | 0.75 | Not reported | Not reported | Not reported | | |
| Lu et al., 2021 [59] | BERT | 0.870 | 0.875 | 0.507 | 0.950 | Not reported | 5-fold cross-validation | Not Reported |
| | SVM | 0.859 | 0.868 | 0.366 | 0.969 | Not reported | | |
| | XGBoost | 0.852 | 0.830 | 0.324 | 0.969 | Not reported | | |
| Moscato et al., 2022 [40] | SVM | 0.73 | 0.71 | 0.90 | 0.52 | Not reported | 10-fold cross-validation | Not Reported |
| | RF | 0.64 | 0.65 | 0.72 | 0.54 | Not reported | | |
| | MP | 0.61 | 0.59 | 0.68 | 0.52 | Not reported | | |
| | LR | 0.65 | 0.66 | 0.73 | 0.49 | Not reported | | |
| | Adaboost | 0.52 | 0.56 | 0.72 | 0.43 | Not reported | | |
| Heintzeman et al., 2013 [60] | NLP | 0.95 | Not reported | Not reported | Not reported | Not reported | Blind test set | Not Reported |
| Miettinen et al., 2021[30] | Cluster | 0.799 | Not reported | 0.741 | 0.877 | Not reported | 1000 Cross-Validation runs | Not Reported |
| | DT | 0.67 | Not reported | Not reported | Not reported | Not reported | | |

| | | | | | | | | |
|---|---------|--------------|--------------|--------------|--------------|--------------|--------------------------|--|
| Akshayaa et al., 2019 [57] | CNN | 0.95 | Not reported | Not reported | Not reported | Not reported | Train/ Test | Not reported |
| Masukawa et al., 2022 [41] | RF | Not reported | 0.90 | Not reported | Not reported | Not reported | 5-fold cross-validation | Not reported |
| | SVM | Not reported | 0.87 | Not reported | Not reported | Not reported | | |
| | LR | Not reported | 0.86 | Not reported | Not reported | Not reported | | |
| | GBM | Not reported | 0.89 | Not reported | Not reported | Not reported | | |
| AI/ML Models used for cancer treatment induced pain research | | | | | | | | |
| Sun et al., 2023 [32] | RF | Not reported | Not reported | 0.362 | 0.914 | Not reported | 10-fold Cross-Validation | XGBoost : 1. ICI; 0.050 (0.038-0.122) 2. E50; 0.046 (0.024-0.123) 3. E90; 0.071 (0.064-0.257) RF: 1.ICI; 0.093 (0.053-0.168) 2. E50; 0.109 (0.034-0.170) 3. E90; 0.129 (0.107-0.303) GBM: 1.ICI; 0.072 (0.049-0.156) 2. E50; 0.071(0.035-0.138) 3. E90; 0.114(0.083-0.286) |
| | LR | Not reported | Not reported | 0.769 | 0.622 | Not reported | | |
| | GBM | Not reported | Not reported | 0.338 | 0.922 | Not reported | | |
| | XGBoost | Not reported | Not reported | 0.339 | 0.907 | Not reported | | |

| | | | | | | | | |
|--------------------------|-----|--------------|--------------|--------------|--------------|--------------|---|--|
| | | | | | | | | Logistic Regression: 1. ICI; 0.070 (0.050 – 0.146) 2. E50; 0.070 (0.027 – 0.127) 3. E90; 0.111 (0.095 – 0.307) |
| Olling et al., 2018 [43] | LR | 0.83 | 0.84 | 0.95 | 0.61 | Not reported | Cross-Validation | Not Reported |
| Juwara et al., 2020 [33] | OLS | Not reported | Not reported | Not reported | Not reported | 1.43 | 10-fold cross-validation | LR: 1. Unadjusted: Intercept: 0.02 (-0.50,0.47) and Slope 0.98 (0.42,1.58) 2. Adjusted: Intercept: 0.03 (-3.4,0.34) and Slope 1.00 (0.40, 1.60) |
| | RR | Not reported | Not reported | Not reported | Not reported | 1.28 | | |
| | EN | Not reported | Not reported | Not reported | Not reported | 1.31 | | |
| | RF | Not reported | Not reported | Not reported | Not reported | 1.39 | | |
| | GBM | Not reported | Not reported | Not reported | Not reported | 1.16 | | |
| | NN | Not reported | Not reported | Not reported | Not reported | 1.50 | | |
| | LR | Not reported | 0.68 | Not reported | Not reported | Not reported | | |
| Sipila et al., 2020 [25] | DT | 0.661 | Not reported | 0.583 | 0.636 | Not reported | Train/test | Not Reported |
| Guan et al., 2023 [26] | RF | Not reported | 0.871 | Not reported | Not reported | Not reported | Train/Test-Internal and external validation | Not Reported |
| | DT | Not reported | 0.864 | Not reported | Not reported | Not reported | | |
| | ANN | Not reported | 0.827 | Not reported | Not reported | Not reported | | |
| | SVM | Not reported | 0.808 | Not reported | Not reported | Not reported | | |

| | | | | | | | | |
|----------------------------|---|--------------|--------------|--------------|--------------|--------------|-------------------------------------|--------------|
| | NB | Not reported | 0.803 | Not reported | Not reported | Not reported | | |
| Lotsch et al., 2017 [23] | Supervised classification ML techniques | 0.944 | Not reported | Not reported | Not reported | Not reported | Not Reported | Not Reported |
| Barber et al., 2022 [35] | RF | Not reported | 0.75 | Not reported | Not reported | Not reported | 5-fold cross-validation | Not Reported |
| Reinbolt et al., 2018 [50] | NAA | Not reported | 0.759 | Not reported | Not reported | Not reported | Leave-One-Out-Cross Validation | Not Reported |
| Wang et al., 2021 [28] | DT | 0.901 | 0.89 | 0.894 | 0.903 | Not reported | Train/test 10-fold cross validation | Not Reported |
| | SVM | Not reported | 0.77 | Not reported | Not reported | Not reported | | |
| | BNN | Not reported | 0.71 | Not reported | Not reported | Not reported | | |
| Lotsch et al., 2018 [36] | Supervised classification model | 0.86 | Not reported | Not reported | Not reported | Not reported | 100-fold cross-validation | Not Reported |
| Sipila et al., 2012 [44] | BM | 0.627 | Not reported | 0.81 | 0.44 | Not reported | Not Reported | Not Reported |
| Lotsch et al., 2018 [36] | Rule Based Classifier | 0.86 | 0.47 | 0.706 | 0.454 | Not reported | Cross-Validation | Not Reported |
| Kringel et al., 2019 [37] | RF | 0.71 | 0.71 | 0.74 | 0.74 | Not reported | Cross-Validation | Not Reported |
| | GBM | 0.71 | 0.71 | 0.69 | 0.74 | Not reported | | |
| | KNN | 0.65 | 0.65 | 0.61 | 0.696 | Not reported | | |
| | NB | 0.67 | 0.67 | 0.696 | 0.696 | Not reported | | |
| | SVM | 0.71 | 0.71 | 0.696 | 0.74 | Not reported | | |
| | LR | 0.65 | 0.65 | 0.57 | 0.74 | Not reported | | |

| | | | | | | | | |
|-------------------------------|------------|--------------|--------------|--------------|--------------|--------------|-------------------------------------|--------------|
| Wang et al., 2021 [16] | DT | 0.878 | 0.652 | 0.017 | 0.167 | Not reported | Train/Test-10-fold Cross validation | Not Reported |
| | RF | 0.871 | 0.666 | 0.050 | 0.222 | ----- | | |
| | XGB | 0.885 | 0.703 | 0.017 | 0.500 | Not reported | | |
| | MLPC | 0.867 | 0.675 | 0.008 | 0.048 | Not reported | | |
| | GNB | 0.721 | 0.685 | 0.500 | 0.205 | Not reported | | |
| | CNN | 0.883 | 0.708 | 0.008 | 0.250 | Not reported | | |
| Kringel et al., 2019 [46] | CART | 0.717 | 0.794 | Not reported | Not reported | Not reported | Train/Test- Cross validation | Not Reported |
| | kNN | 0.739 | 0.739 | Not reported | Not reported | Not reported | | |
| | SVM | 0.804 | 0.826 | Not reported | Not reported | Not reported | | |
| | Regression | 0.739 | 0.807 | Not reported | Not reported | Not reported | | |
| | NB | 0.739 | 0.802 | Not reported | Not reported | Not reported | | |
| Lotsch et al., 2022 [53] | RF | 0.58 | 0.58 | 0.65 | 0.50 | Not reported | Cross-Validation | Not Reported |
| Kumar et al., 2023 [34] | GLM | Not reported | Not reported | Not reported | Not reported | 0.130 | 10-fold cross-validation | Not Reported |
| | SVM | Not reported | Not reported | Not reported | Not reported | 0.127 | | |
| | RF | Not reported | Not reported | Not reported | Not reported | 0.129 | | |
| | NN | Not reported | Not reported | Not reported | Not reported | 0.132 | | |
| Facciorusso et al., 2019 [51] | ANN | Not reported | 0.94 | Not reported | Not reported | 0.057 | 10-fold Cross-Validation | Not Reported |
| | LR | Not reported | 0.85 | Not reported | Not reported | 0.147 | | |
| Dolendo et al., 2022 [45] | LR | Not reported | 0.763 | Not reported | Not reported | Not reported | 10-Fold Cross- | Not Reported |

| | | | | | | | |
|--|------------------------|--------------|-------|--------------|--------------|--------------|------------|
| | Ridge Regression | Not reported | 0.775 | Not reported | Not reported | Not reported | Validation |
| | Lasso | Not reported | 0.799 | Not reported | Not reported | Not reported | |
| | Elastic Net Regression | Not reported | 0.801 | Not reported | Not reported | Not reported | |

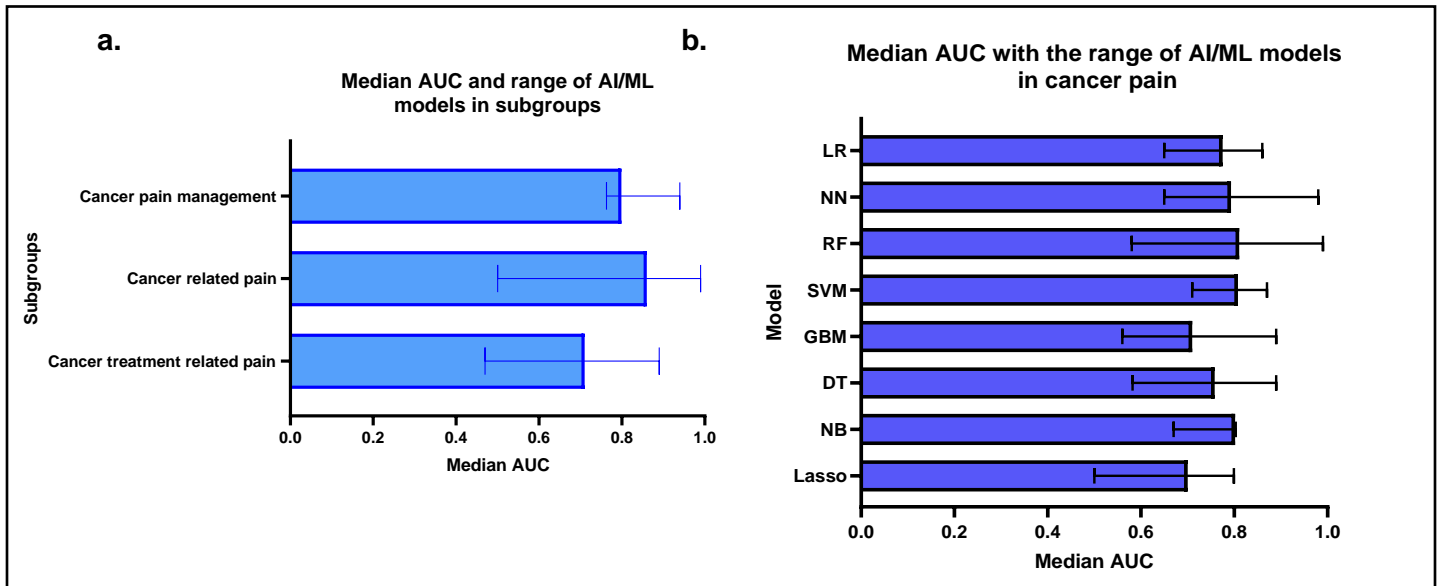


Figure 5: a. Median area under the receiver operating curve (AUC) across all included studies by subgroups. **b.** Median area under the receiver operating curve (AUC) across all included studies by AI/ML model.

Adherence to TRIPOD Guidelines

The overall compliance with the TRIPOD guidelines reporting checklist was 70.7%, revealing that 7 out of 31 domains fell below a 60% adherence rate. While reporting adherence surpassed 80% for components such as Study Design, Eligibility Criteria and Statistical Methods, it plummeted below 50% for crucial elements like blinding of outcomes/predictors, handling of missing data, model development, and identification of risk groups. Specifically, 75.4% of studies adequately specified their study population

concerning inclusion/exclusion criteria and baseline characteristics. Lastly, 95.5% of abstracts provided adequate information on study methodology, and approximately 70% of studies disclosed funding sources (Figure 6).

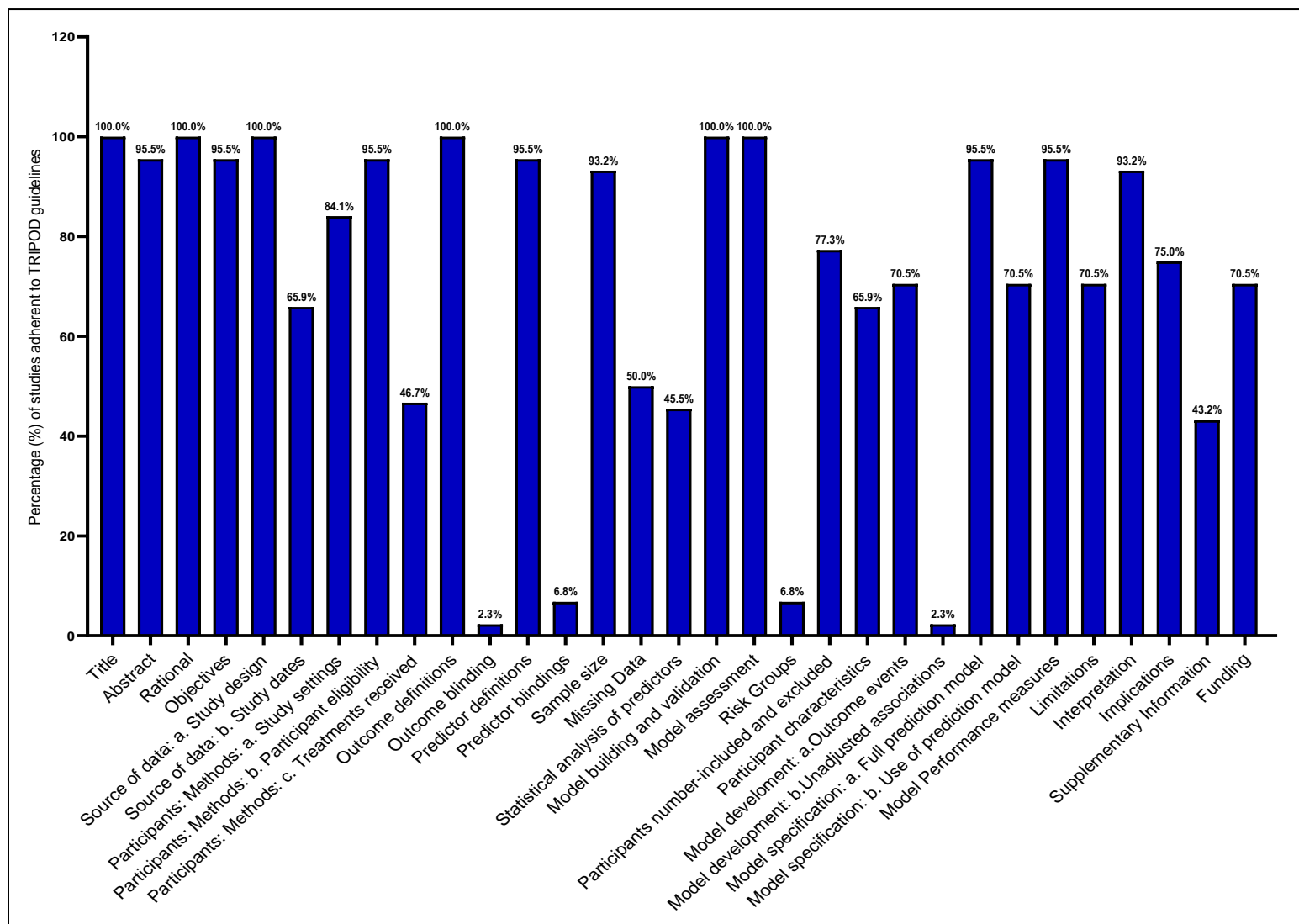


Figure 6: Frequency of adherence on included studies to reporting checklist of Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) guidelines.

Discussion:

This comprehensive review delves into 44 studies that were identified through database searches, each utilizing distinct AI/ML methodologies within the field of cancer pain research published between 2006 and 2023. Most studies used prospective uni-institutional cohorts of cancer patients. In comparison to conventional statistical methods, AI/ML techniques have demonstrated superior performance, particularly in cancer pain prediction and pain management according to Sun et al. (2023) [32] and Juwara et al. (2020) [33]. The utilization of ML models, trained on big data sets of multiple features, has exhibited impressive mapping to select categories, classify patients and guide pain management decisions.

The results in this systematic review highlighted the substantial benefits of implementing AI/ML in cancer pain research, particularly in the context of predicting pain in cancer patients who have undergone cancer treatments. Out of the 44 studies, 19 focused on post-cancer treatment pain prediction, and 18 studies aimed at enhancing pain prediction and identifying high-risk factors associated with pain in cancer patients. Furthermore, our review pinpointed 7 studies that applied AI/ML for predicting cancer pain management and aiding in decisions regarding pain management (e.g., opioids usage).

Our comprehensive review illuminated the diverse array of AI/ML models deployed in the field of cancer pain research. Most of these models took the form of supervised classification or regression algorithms, including DT, RF, GBM, and LR. In addition, unsupervised clustering algorithms were applied to classify subgroups of pain patients and to identify key pain-related parameters. For making informed pain management decisions, advanced NNs were utilized and validated. AI-based decision support systems emerged as promising tools in guiding pain management choices. Most studies investigated multiple models (each model was tested separately) (55%), where RF and LR were the most common models used in these multiple models' studies (n=15 studies for each). RF models demonstrated the highest performance across all studies (median AUC 81%), and Lasso models demonstrated the highest sensitivity (100%) while the lowest specificity (0%).

The expansion of clinical, genetic, and healthcare-related data has created an impetus for leveraging cutting-edge AI/ML techniques to harness this wealth of big data for the betterment of healthcare outcomes, especially for cancer patients contending with both the disease and its treatments. In our review, studies collectively demonstrated the potential of AI/ML models in predicting, assessing, and understanding pain among cancer patients, utilizing various inputs such as patient characteristics, imaging, video analysis, and clinical variables to improve pain management and remote consultations. While most studies used demographic and clinical data as input variables, only 4 studies included genetic data as inputs, Olesen et al., and Reinbolt et al., used single nucleotide polymorphisms (SNPs) data as inputs for the models [47, 50]. More novel approaches were observed with studies including image data (pathology, radiological, dosiomic data) as model inputs related to cancer pain (Akshayaa et al.[57], and Chao et al., [24]). Our

systematic review did not provide specific details on the evaluation process of these features or how they were selected for model input. Deep learning offers the potential of analyzing high dimensional data (e.g., images) in combination with non-image data. This is an opportunity that future studies should take towards more comprehensive models in cancer pain prediction. Five studies used video recording, machine signals or tests results as input data of the models [23, 34, 40, 52, 54]. Text mining and NLP open the opportunities for future studies in cancer pain using texts and different non-structured data,

Numerous studies have explored the power of AI and ML to predict pain in various medical diseases. For instance, Matsangidou et al., (2021) conducted a comprehensive investigation into the utilization of AI and ML techniques for pain prediction, showcasing the potential of these advanced technologies to enhance the accuracy and effectiveness of pain prediction and prognosis [65]. Such research exemplifies the growing significance of AI and ML in improving patient care and healthcare outcomes. However, studies specifically addressing cancer-related pain are limited, and there is a lack of data regarding the calibration of models and adherence to TRIPOD guidelines. To our knowledge, this is the first study to do a comprehensive investigation and analysis of all up-to-date studies focused on applying AI/ML models in cancer pain medicine and in cancer pain management decisions. We analyzed the different models applied in these studies and the median model performance. We categorized the studies according to the use of the models into three subgroups: cancer pain research, cancer treatment related pain, and cancer treatment (e.g., opioids) decisions. Our data analysis demonstrated an increase in the trend of AI/ML studies in cancer pain in the last few years, and that AI/ML models showed high performance in cancer pain prediction, classification and management decisions.

In our review, we extracted and analyzed data regarding the adherence to The Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines, models calibration, external validation, and clinical application of the AI/ML models. Interestingly, our results revealed 70.7% overall compliance with the TRIPOD reporting checklist. However, few studies tested models' calibration (5%), performed external validation of internally tested models (n=6, 14%) or discussed the clinical application of the validated models (23%). According to Van Calster et. al (2019) [66], the performance assessment of AI/ML models that estimate disease risks or predict health outcomes for clinical decision-making should involve evaluating model discrimination (e.g., calculating ROC-AUC) and model calibration as essential elements in the evaluation process [66]. While most studies on AI/ML models focus on discrimination and classification performance, calibration of models is often overlooked. Poor calibration of predictive algorithms can be misleading, leading to incorrect and potentially harmful clinical decisions [66]. Furthermore, adhering to the TRIPOD guidelines is essential in the context of AI/ML models, these guidelines provide a structured framework for transparent and comprehensive reporting, ensuring the reliability and reproducibility of predictive models [17, 18]. Equally crucial is the process

of external validation, where models are rigorously tested in diverse and independent datasets to assess their generalizability and reliability beyond the initial training data. This step is vital in affirming the robustness of the models and their applicability to real-world scenarios [67]. Additionally, recurrent local validation should be considered in addition to external validation to test models' reliability, safety and generalizability for clinical application, and Youssef et al., (2023) proposed the Machine Learning Operations-inspired paradigm for recurrent local validation of AI/ML models to maintain the validity of the models [68]. Furthermore, the clinical application of established ML models is of utmost importance to bridge the gap between research and practical healthcare settings. Understanding how these models perform in clinical practice enhances their utility and ensures informed decision-making. Several studies investigate robust AI/ML models in healthcare, however, few of these models have been clinically applied due to several limitations to translate AI/ML into clinical practices [69]. By prioritizing adherence to TRIPOD guidelines, conducting thorough repetitive local validations, external validations, and emphasizing the clinical application of ML models, the healthcare community can foster trust in AI-based tools, ultimately leading to improved patient outcomes and more effective healthcare interventions.

While deep reinforcement learning (RL) has exhibited remarkable efficacy in making morphine dosage decisions and optimizing pain management within the intensive care unit (ICU), as demonstrated by Lopez-Martinez et al. (2019) [70], our review, unfortunately, did not uncover any studies that had leveraged RL for the optimization of cancer pain management. This observation underscores a potential avenue for future research and development in the field, highlighting the need for exploring the application of RL techniques to enhance the management of pain in cancer patients.

Studies included in our review collectively demonstrate diverse applications of AI/ML models in cancer pain research, illuminating their potential in predicting, understanding patient satisfaction, identifying pain-related attributes, and managing pain in cancer patients across different age groups and healthcare settings. The identification of high-risk cancer pain patients and their associated risk factors is instrumental in stratifying patients according to their pain risk, which, in turn, informs the development of personalized pain management strategies. The integration of AI and ML in cancer pain prediction and patient risk stratification streamlines the analysis of complex big data with minimal human intervention, offering a promising path to improved outcomes and enhanced QOL for cancer patients suffering with pain. Furthermore, AI/ML techniques have proven effective in not only predicting but also diagnosing, categorizing, and recommending treatments for cancer pain. Although the clinical importance of applying these AI/ML models, several studies included in our review provided very limited details on the clinical validation of their AI/ML models in real healthcare settings, which is a crucial step in ensuring the applicability and reliability of these models in clinical practice. Clinical validation, in larger and more diverse cohorts, is vital to establish the predictive value of identified features and optimize pain treatment for cancer patients, enhancing the model's real-world applicability, generalizability and effectiveness in clinical settings.

Limitations:

While our diligent effort involved an extensive comprehensive exploration of the implementation of AI and ML in the domain of cancer pain medicine, it is important to have some caution in interpreting the results due to the limitations in the design of our study. Notably, we encountered substantial heterogeneity among the identified studies, encompassing variations in the models employed, diverse AI/ML performance metrics, and disparities in the outputs associated with pain. Additionally, there is a lack of some AI/ML performance metrics which may affect the overall models' performance. Furthermore, it's essential to acknowledge that our search strategies did not explore other symptoms or cancer therapy toxicities that may have relevance to pain, such as mucositis, dysphagia, pneumonitis, and more. Our primary focus predominantly centered on patient-reported pain phenotypes, a deliberate choice that should be considered when evaluating the scope and implications of our findings. Very few data were available on the model calibration, external validation, and clinical application of the tested models in included articles which adds limitation of interpretation of the degree of biases, robustness, and clinical reliability of the applied models.

Future Directions:

Generalizability of the performance of the identified models that demonstrated high performance is required with external validation. Assessment of the robustness, clinical application of these models is needed to be conducted in the future for clinical use in real-world healthcare settings. The use of DL and RL models in the inclusion of imaging data for cancer pain medicine should be further explored.

Conclusion:

Recent advancements in AI and ML techniques have ushered in a new era of cancer pain research, with applications including cancer pain prediction, the anticipation of pain induced by cancer treatments, and aiding in pain management decisions. AI/ML models, showed high performance in predicting cancer pain, risk stratification, and enabling personalized pain management, with a median AUC 0.77 across all models and all studies. Although the adherence to TRIPOD guidelines was 70.7%, testing models' calibration was 5%, the models' external validation was 14% and the clinical application was 23%. There is an ongoing need for AI/ML models to be rigorously tested for calibration and externally clinically validated before their implementation in the real-world healthcare setting.

Funding Statement:

Drs. Moreno and Fuller received project related grant support from the National Institutes of Health (NIH)/ National Institute of Dental and Craniofacial Research (NIDCR) (Grants R21DE031082); Dr. Moreno received salary support from NIDCR (K01DE03052) and the National Cancer Institute (K12CA088084) during the project period. Dr. Fuller receives grant and infrastructure support from MD Anderson Cancer Center via: the Charles and Daneen Stiefel Center for Head and Neck Cancer Oropharyngeal Cancer Research Program; the Program in Image-guided Cancer Therapy; and the NIH/NCI Cancer Center Support Grant (CCSG) Radiation Oncology and Cancer Imaging Program (P30CA016672). Dr. Fuller has received unrelated direct industry grant/in-kind support, honoraria, and travel funding from Elekta AB. Vivian Salama was funded by Dr. Moreno's fund supported from Paul Calabresi K12CA088084 Scholars Program. Kareem Wahid was supported by an Image Guided Cancer Therapy (IGCT) T32 Training Program Fellowship from T32CA261856".

Conflict of interest:

Authors declare that they have no known competing commercial, financial interests or personal relationships that could be constructed as potential conflict of interest.

Acknowledgement:

We thank The American Legion Auxiliary (ALA) for the ALA Fellowship in Cancer Research, 2022-2024, through The University of Texas, MD Anderson Cancer Center, UTHealth Houston Graduate School of Biomedical Sciences (GSBS). We thank The McWilliams School of Biomedical Informatics at UTHealth Houston, The Student Governance Organization (SGO) Student Excellence award.

References:

1. Merskey H, Bogduk N, International Association for the Study of Pain. Task Force on Taxonomy. Classification of chronic pain : descriptions of chronic pain syndromes and definitions of pain terms. 2nd ed. Seattle: IASP Press; 1994. xvi, 222 p. p.
2. Swarm RA, Paice JA, Anghelescu DL, Are M, Bruce JY, Buga S, Chwistek M, Cleeland C, Craig D, Gafford E, Greenlee H, Hansen E, Kamal AH, Kamdar MM, LeGrand S, Mackey S, McDowell MR, Moryl N, Nabell LM, Nesbit S, Bcps, O'Connor N, Rabow MW, Rickerson E, Shatsky R, Sindt J, Urba SG, Youngwerth JM, Hammond LJ, Gurski LA. Adult Cancer Pain, Version 3.2019, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw*. 2019;17(8):977-1007.
3. Institute of Medicine (IOM). Relieving pain in America: a blueprint for transforming prevention, care, education, and research. Washington, DC: National Academies Press; 2011. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK91497/>. (Accessed January 4, 2019) 2019 [
4. Paice JA, Portenoy R, Lacchetti C, Campbell T, Cheville A, Citron M, Constine LS, Cooper A, Glare P, Keefe F, Koyyalagunta L, Levy M, Miaskowski C, Otis-Green S, Sloan P, Bruera E.

Management of Chronic Pain in Survivors of Adult Cancers: American Society of Clinical Oncology Clinical Practice Guideline. *J Clin Oncol*. 2016;34(27):3325-45.

5. Paice JA. Cancer Pain Management: Strategies for Safe and Effective Opioid Prescribing. *J Natl Compr Canc Netw*. 2016;14(5 Suppl):695-7.
6. Hjermland MJ, Kaasa S, Caraceni A, Loge JH, Pedersen T, Haugen DF, Aass N, European Palliative Care Research C. Characteristics of breakthrough cancer pain and its influence on quality of life in an international cohort of patients with cancer. *BMJ Support Palliat Care*. 2016;6(3):344-52.
7. Portenoy RK, Hagen NA. Breakthrough pain: definition, prevalence and characteristics. *Pain*. 1990;41(3):273-81.
8. Grossman SA NS. Cancer-related Pain. In: Niederhuber JE, Armitage JO, Doroshow JH, Kastan MB, Tepper JE, eds. *Abeloff's Clinical Oncology*. 5th ed. Philadelphia: Elsevier, 2014:608-619. 2014.
9. American Cancer Society. Acute, Chronic, and Breakthrough Pain. Cancer.org. chrome-extension://efaidnbnmnnibpcajpcgclclefindmkaj/<https://www.cancer.org/content/dam/CRC/PDF/Public/9379.00.pdf> [
10. Cascella M, Coluccia S, Monaco F, Schiavo D, Nocerino D, Grizzuti M, Romano MC, Cuomo A. Different Machine Learning Approaches for Implementing Telehealth-Based Cancer Pain Management Strategies. *J Clin Med*. 2022;11(18).
11. WHO's cancer pain ladder for adults [Internet].
12. Char DS, Shah NH, Magnus D. Implementing Machine Learning in Health Care - Addressing Ethical Challenges. *N Engl J Med*. 2018;378(11):981-3.
13. Sendak M, Gao M, Nichols M, Lin A, Balu S. Machine Learning in Health Care: A Critical Appraisal of Challenges and Opportunities. *EGEMS (Wash DC)*. 2019;7(1):1.
14. Mitchell TM. *Machine Learning* 1997.
15. Mitchell TM. Does machine learning really work? *AI Magazine*. 18(3):11.
16. Wang Y, Zhu Y, Xue Q, Ji M, Tong J, Yang JJ, Zhou CM. Predicting chronic pain in postoperative breast cancer patients with multiple machine learning and deep learning models. *J Clin Anesth*. 2021;74:110423.
17. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, Logullo P, Beam AL, Peng L, Van Calster B, van Smeden M, Riley RD, Moons KG. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open*. 2021;11(7):e048008.
18. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD). *Ann Intern Med*. 2015;162(10):735-6.
19. Vaicenavicius J, Widmann D, Andersson C, Lindsten F, Roll J, Schön TB. Evaluating model calibration in classification. *Pr Mach Learn Res*. 2019;89.
20. Roberta Pappadà FP. Discrimination in machine learning algorithms. arXiv:220700108 [statML]. 2022.
21. Covidence. Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia. Available at www.covidence.org. [
22. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097.
23. Lotsch J, Ultsch A, Kalso E. Prediction of persistent post-surgery pain by preoperative cold pain sensitivity: biomarker development with machine-learning-derived analysis. *Br J Anaesth*. 2017;119(4):821-9.
24. Chao HH, Valdes G, Luna JM, Heskell M, Berman AT, Solberg TD, Simone CB, 2nd. Exploratory analysis using machine learning to predict for chest wall pain in patients with stage I non-small-cell lung cancer treated with stereotactic body radiation therapy. *J Appl Clin Med Phys*. 2018;19(5):539-46.

25. Sipila R, Kalso E, Lotsch J. Machine-learned identification of psychological subgroups with relation to pain interference in patients after breast cancer treatments. *Breast*. 2020;50:71-80.
26. Guan Y, Tian Y, Fan YW. Pain management in patients with hepatocellular carcinoma after transcatheter arterial chemoembolisation: A retrospective study. *World J Gastrointest Surg*. 2023;15(3):374-86.
27. Bobrova O, Zyryanov S, Shnayder N, Petrova M. Personalized Calculator for Prediction of Opioid-Associated Pharmacoresistance in Patients with Pan Crease Cancer. *Arch Euromedica*. 2020;10(4):20-2.
28. Wang Z, Sun J, Sun Y, Gu Y, Xu Y, Zhao B, Yang M, Yao G, Zhou Y, Li Y, Du D, Zhao H. Machine Learning Algorithm Guiding Local Treatment Decisions to Reduce Pain for Lung Cancer Patients with Bone Metastases, a Prospective Cohort Study. *Pain Ther*. 2021;10(1):619-33.
29. Shimada K, Tsuneto S. Novel method for predicting nonvisible symptoms using machine learning in cancer palliative care. *Sci Rep*. 2023;13(1):12088.
30. Miettinen T, Mantyselka P, Hagelberg N, Mustola S, Kalso E, Lotsch J. Machine learning suggests sleep as a core factor in chronic pain. *Pain*. 2021;162(1):109-23.
31. Pombo N, Rebelo P, Araujo P, Viana J. Design and evaluation of a decision support system for pain management based on data imputation and statistical models. *Measurement*. 2016;93:480-9.
32. Sun C, Li M, Lan L, Pei L, Zhang Y, Tan G, Zhang Z, Huang Y. Prediction models for chronic postsurgical pain in patients with breast cancer based on machine learning approaches. *Front Oncol*. 2023;13:1096468.
33. Juwara L, Arora N, Gornitsky M, Saha-Chaudhuri P, Velly AM. Identifying predictive factors for neuropathic pain after breast cancer surgery using machine learning. *Int J Med Inform*. 2020;141:104170.
34. Kumar S, Kesavan R, Sistla SC, Penumadu P, Natarajan H, Chakradhara Rao US, Nair S, Vasuki V, Kundra P. Predictive models for fentanyl dose requirement and postoperative pain using clinical and genetic factors in patients undergoing major breast surgery. *Pain*. 2023;164(6):1332-9.
35. Barber EL, Garg R, Strohl A, Roque D, Tanner E. Feasibility and Prediction of Adverse Events in a Postoperative Monitoring Program of Patient-Reported Outcomes and a Wearable Device Among Gynecologic Oncology Patients. *JCO Clin Cancer Inform*. 2022;6(1):e2100167.
36. Lotsch J, Sipila R, Tasmuth T, Kringel D, Estlander AM, Meretoja T, Kalso E, Ultsch A. Machine-learning-derived classifier predicts absence of persistent pain after breast cancer surgery with high accuracy. *Breast Cancer Res Treat*. 2018;171(2):399-411.
37. Kringel D, Kaunisto MA, Kalso E, Lotsch J. Machine-learned analysis of the association of next-generation sequencing-based genotypes with persistent pain after breast cancer surgery. *Pain*. 2019;160(10):2263-77.
38. Cascella M, Scarpati G, Bignami EG, Cuomo A, Vittori A, Di Gennaro P, Crispo A, Coluccia S. Utilizing an artificial intelligence framework (conditional generative adversarial network) to enhance telemedicine strategies for cancer pain management. *J Anesth Analg Crit Care*. 2023;3(1):19.
39. Lou Z, Vivas-Valencia C, Shields CG, Kong N. Examining how physician factors influence patient satisfaction during clinical consultations about cancer prognosis and pain. *PEC Innov*. 2022;1:100017.
40. Moscato S, Orlandi S, Giannelli A, Ostan R, Chiari L. Automatic pain assessment on cancer patients using physiological signals recorded in real-world contexts. *Annu Int Conf IEEE Eng Med Biol Soc*. 2022;2022:1931-4.
41. Masukawa K, Aoyama M, Yokota S, Nakamura J, Ishida R, Nakayama M, Miyashita M. Machine learning models to detect social distress, spiritual pain, and severe physical psychological symptoms in terminally ill patients with cancer from unstructured text data in electronic medical records. *Palliative Med*. 2022;36(8):1207-16.

42. Knudsen AK, Brunelli C, Kaasa S, Apolone G, Corli O, Montanari M, Fainsinger R, Aass N, Fayers P, Caraceni A, Klepstad P, EPCRC, EPO EPS. Which variables are associated with pain intensity and treatment response in advanced cancer patients? - Implications for a future classification system for cancer pain. *Eur J Pain*. 2011;15(3):320-7.
43. Olling K, Nyeng DW, Wee L. Predicting acute odynophagia during lung cancer radiotherapy using observations derived from patient-centred nursing care. *Tech Innov Patient Support Radiat Oncol*. 2018;5:16-20.
44. Sipila R, Estlander AM, Tasmuth T, Kataja M, Kalso E. Development of a screening instrument for risk factors of persistent pain after breast cancer surgery. *Br J Cancer*. 2012;107(9):1459-66.
45. Dolendo IM, Wallace AM, Armani A, Waterman RS, Said ET, Gabriel RA. Predictive Analytics for Inpatient Postoperative Opioid Use in Patients Undergoing Mastectomy. *Cureus*. 2022;14(3):e23079.
46. Kringel D, Kaunisto MA, Kalso E, Lotsch J. Machine-learned analysis of global and glial/opioid intersection-related DNA methylation in patients with persistent pain after breast cancer surgery. *Clin Epigenetics*. 2019;11(1):167.
47. Olesen AE, Gronlund D, Gram M, Skorpen F, Drewes AM, Klepstad P. Prediction of opioid dose in cancer pain patients using genetic profiling: not yet an option with support vector machine learning. *BMC Res Notes*. 2018;11(1):78.
48. Lotsch J, Ultsch A, Kalso E. Data-science-based subgroup analysis of persistent pain during 3 years after breast cancer surgery: A prospective cohort study. *Eur J Anaesthesiol*. 2020;37(3):235-46.
49. Pantano F, Manca P, Armento G, Zeppola T, Onorato A, Iuliani M, Simonetti S, Vincenzi B, Santini D, Mercadante S, Marchetti P, Cuomo A, Caraceni A, Mediati RD, Vellucci R, Mammucari M, Natoli S, Lazzari M, Dauri M, Adile C, Airoldi M, Azzarello G, Blasi L, Chiurazzi B, Degiovanni D, Fusco F, Guardamagna V, Liguori S, Palermo L, Mamei S, Masedu F, Mazzei T, Melotti RM, Menardo V, Miotti D, Moroso S, Pascoletti G, De Santis S, Orsetti R, Papa A, Ricci S, Scelzi E, Sofia M, Aielli F, Valle A, Tonini G. Breakthrough Cancer Pain Clinical Features and Differential Opioids Response: A Machine Learning Approach in Patients With Cancer From the IOPS-MS Study. *JCO Precis Oncol*. 2020;4.
50. Reinbolt RE, Sonis S, Timmers CD, Fernandez-Martinez JL, Cernea A, de Andres-Galiana EJ, Hashemi S, Miller K, Pilarski R, Lustberg MB. Genomic risk prediction of aromatase inhibitor-related arthralgia in patients with breast cancer using a novel machine-learning algorithm. *Cancer Med*. 2018;7(1):240-53.
51. Facciorusso A, Del Prete V, Antonino M, Buccino VR, Muscatiello N. Response to repeat echoendoscopic celiac plexus neurolysis in pancreatic cancer patients: A machine learning approach. *Pancreatology*. 2019;19(6):866-72.
52. Cascella M, Vitale VN, Mariani F, Iuorio M, Cutugno F. Development of a binary classifier model from extended facial codes toward video-based pain recognition in cancer patients. *Scand J Pain*. 2023.
53. Lotsch J, Mustonen L, Harno H, Kalso E. Machine-Learning Analysis of Serum Proteomics in Neuropathic Pain after Nerve Injury in Breast Cancer Surgery Points at Chemokine Signaling via SIRT2 Regulation. *Int J Mol Sci*. 2022;23(7).
54. Sokouti B, Haghypour S. Pain Management Based on Spinal Cord Dorsal Horn System Response Identification Using Artificial Neural Networks. *Biomed Eng-App Bas C*. 2014;26(3).
55. Xuyi W, Seow H, Sutradhar R. Artificial neural networks for simultaneously predicting the risk of multiple co-occurring symptoms among patients with cancer. *Cancer Med*. 2021;10(3):989-98.
56. Bang YH, Choi YH, Park M, Shin SY, Kim SJ. Clinical relevance of deep learning models in predicting the onset timing of cancer pain exacerbation. *Sci Rep*. 2023;13(1):11501.

57. Akshayaa S, Vidhya R, Vyshnavi AMH, Namboori PKK. Exploring Pain Insensitivity Inducing Gene ZFX2 by using Deep Convolutional Neural Network. Proceedings of the 2019 3rd International Conference on Computing Methodologies and Communication (Iccmc 2019). 2019:68-72.
58. DiMartino L, Miano T, Wessell K, Bohac B, Hanson LC. Identification of Uncontrolled Symptoms in Cancer Patients Using Natural Language Processing. *J Pain Symptom Manage*. 2022;63(4):610-7.
59. Lu Z, Sim JA, Wang JX, Forrest CB, Krull KR, Srivastava D, Hudson MM, Robison LL, Baker JN, Huang IC. Natural Language Processing and Machine Learning Methods to Characterize Unstructured Patient-Reported Outcomes: Validation Study. *J Med Internet Res*. 2021;23(11):e26777.
60. Heintzelman NH, Taylor RJ, Simonsen L, Lustig R, Anderko D, Haythornthwaite JA, Childs LC, Bova GS. Longitudinal analysis of pain in patients with metastatic prostate cancer using natural language processing of medical record text. *J Am Med Inform Assoc*. 2013;20(5):898-905.
61. Im EO, Chee W. Evaluation of the decision support computer program for cancer pain management. *Oncol Nurs Forum*. 2006;33(5):977-82.
62. Im EO, Chee W. The DSCP-CA: a decision support computer program--cancer pain management. *Comput Inform Nurs*. 2011;29(5):289-96.
63. Xu X, Yang K, Zhang FL, Liu WW, Wang YY, Yu CY, Wang JY, Zhang KK, Zhang C, Nenadic G, Tao DC, Zhou XZ, Shang HC, Chen JX. Identification of herbal categories active in pain disorder subtypes by machine learning help reveal novel molecular mechanisms of algesia. *Pharmacol Res*. 2020;156.
64. Lotsch J, Sipila R, Dimova V, Kalso E. Machine-learned selection of psychological questionnaire items relevant to the development of persistent pain after breast cancer surgery. *Br J Anaesth*. 2018;121(5):1123-32.
65. Matsangidou M, Liampas A, Pittara M, Pattichi CS, Zis P. Machine Learning in Pain Medicine: An Up-To-Date Systematic Review. *Pain Ther*. 2021;10(2):1067-84.
66. van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Initiative S. Calibration: the Achilles heel of predictive analytics. *Bmc Med*. 2019;17(1).
67. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J*. 2021;14(1):49-58.
68. Youssef A, Pencina M, Thakur A, Zhu T, Clifton D, Shah NH. External validation of AI models in health should be replaced with recurring local validation. *Nat Med*. 2023;29(11):2686-7.
69. Busnatu S, Niculescu AG, Bolocan A, Petrescu GED, Paduraru DN, Nastasa I, Lupusoru M, Geanta M, Andronic O, Grumezescu AM, Martins H. Clinical Applications of Artificial Intelligence-An Updated Overview. *J Clin Med*. 2022;11(8).
70. Lopez-Martinez D, Eschenfeldt P, Ostvar S, Ingram M, Hur C, Picard R. Deep Reinforcement Learning for Optimal Critical Care Pain Management with Morphine using Dueling Double-Deep Q Networks. *Annu Int Conf IEEE Eng Med Biol Soc*. 2019;2019:3960-3.