Artificial Intelligence

The Analysis of Pain Research through the Lens of Artificial Intelligence and Machine Learning

Jagadesh N. Nagireddi, MEng[1], Amanya Ketan Vyas[1], Mahendra R. Sanapati, MD[1], Amol Soin, MD[2,3], and Laxmaiah Manchikanti, MD[1]

From: [1]Pain Management Centers of America, Paducah, KY & Evansville, IN; [2]Ohio Pain Clinic, Centerville, OH; [3]Wright State University, Dayton, OH

See page E239 for more affiliation information.

Address Correspondence:
Laxmaiah Manchikanti, MD
Pain Management Centers of America
67 Lakeview Drive
Paducah, Kentucky 42001
E-mail: drlm@thepainmd.com

Disclaimer: There was no external funding in the preparation of this manuscript.

Conflict of interest: Dr. Soin is the founder and CEO of Soin Neuroscience, which is developing a spinal cord stimulator to treat spinal pain and has a patent for Soin Neuroscience, Jan One, and Avanos and a patent pending for Soin Therapeutics. All other authors certify that he or she, or a member of his or her immediate family, has no commercial association (i.e., consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) that might pose a conflict of interest in connection with the submitted manuscript.

Manuscript received: 11-20-2021
Revised manuscript received: 01-08-2022
Accepted for publication: 01-29-2022

Free full manuscript: www.painphysicianjournal.com

Background: Traditional pain assessment methods have significant limitations due to the high variability in patient reported pain scores and perception of pain by different individuals. There is a need for generalized and automatic pain detection and recognition methods. In this paper, state-of-the-art machine learning (ML) and deep learning methods in this field are analyzed as well as pain management techniques.

Objective: The objective of the study is to analyze the current use of artificial intelligence (AI) and ML in the analysis and management of pain and to disseminate this knowledge prompting future utilization by medical professionals.

Study Design: A narrative review of the literature focusing on the latest algorithms in AI and ML for pain assessment and management.

Methods: Research studies were collected using a literature search on PubMed, Science Direct and IEEE Xplore between 2018 and 2020.

Results: The results of our assessment resulted in the identification of 47 studies meeting inclusion criteria. Pain assessment was the most studied subject with 11 studies, followed by automated measurements with 10 studies, spinal diagnosis with 8 studies, facial expression with 7 studies, pain assessment in special settings evaluated in 5 studies, 4 studies described treatment algorithms, and 2 studies assessed neonatal pain. These studies varied from simple to highly complex methodology. The majority of the studies suffered from inclusion of a small number of patients and without replication of results. However, considering AI and ML are dynamic and emerging specialties, the results shown here are promising. Consequently, we have described all the available literature in summary formats with commentary. Among the various assessments, facial expression and spinal diagnosis and management appear to be ready for inclusion as we continue to progress.

Limitations: This review is not a systematic review of ML and AI applications in pain research. This review only provides a general idea of the upcoming techniques but does not provide an authoritative evidence-based conclusive opinion of their clinical application and effectiveness.

Conclusion: While a majority of the studies focused on classification tasks, very few studies have explored the diagnosis and management of pain. Usage of ML techniques as support tools for clinicians holds an immense potential in the field of pain management.

Key words: Pain assessment, pain management, pain prediction, machine learning, deep learning, artificial intelligence, numeric rating scale, facial pain, chronic pain

Pain Physician 2022: 25:E211-E243

Pain is a vital feedback mechanism in the human body that functions to keep the body in homeostasis. It signals the body in response to damage and aids in the prevention of further illness. Although pain is a universal sensation, "pain and pain chronification are incompletely understood and unresolved medical problems" (1). Chronic pain is the most prevalent chronic disease across the globe, negatively impacting the quality of life (QOL) and function, impacting individuals, their families, communities, and health systems. Chronic pain is the leading cause of disability straining not only the healthcare system, but the family unit. Overall, the impact of chronic pain, of which spinal pain is the major component with low back pain as the leading cause, continues to be disproportionate and enormous (2-14). Assessments of the impact of spinal pain in the US have shown low back pain to rank number 1, neck pain ranking number 3, with musculoskeletal disorders ranking number 2, depression and anxiety placed in the 4th and 5th places, among the 30 leading diseases and injuries contributing to years lived with disability in 2010 (12). In reference to the economic impact of health care in the US, Dieleman et al (13,14) showed an estimated spending of $134.5 billion in 2016, a 53.5% increase from 2013, or $87.6 billion spent for managing spinal pain. The costs of other musculoskeletal disorders also increased by 43.5% from $183 billion in 2013 to $263.3 billion in 2016. In addition, national health expenditures (15) are projected to grow at an average annual rate of 5.4% from 2019 to 2028 and to represent 19.7% of the gross domestic product (GDP) by the end of the period in the US. Further, healthcare expenditures have been escalating and the financial impact on the US economy is growing with a perfect storm created by COVID-19, the opioid epidemic, issues related to regulations, and lack of reliable, unbiased, evidence-based medicine (4,16-20).

Chronic pain is defined by the International Association for the Study of Pain (IASP) as, "pain that exists beyond an expected timeframe of healing (21)." Further, a descriptive definition provided by American Society of Interventional Pain Physicians (ASIPP) defined chronic pain as, "pain that persists 6 months after an injury and beyond the usual course of an acute disease or a reasonable time for a comparable injury to heal, that is associated with chronic pathologic processes that cause continuous or intermittent pain for months or years that may continue in the presence or absence of demonstrable pathologies; may not be amenable to routine pain control methods; and healing may never occur (22,23)."

However, the true burden lies in the proper assessment of pain and outcomes with numerous modalities of treatments utilized (1-4). Currently, the gold standard of pain measurement is self-reporting through instruments including the Visual Analogue Scale (VAS) and the Numeric Rating Scale (NRS), along with multiple behavioral scales. These methods of evaluation are subject to high variability in individual perceptions of pain (1). Pain can either be underreported or overreported based on a confluence of physical, psychological, motivational, and other factors. Similar to reports of pain assessment, multiple instruments have been published to assess the functional status of a person suffering with chronic pain and assessment of outcomes including Oswestry Disability Index (ODI) (24,25), Roland-Morris Disability Questionnaire (RMDQ) (25), Quebec Back Pain Disability Scales (QBPDS) (25), Neck Disability Index (NDI) (26), EuroQOL Five Dimensional Questionnaire (EQ-5D-3L) (27), Pain Disability Index (PDI) (28), Global Mental Health (GMH), and Global Physical Health (GPH) scales, and many other tests (29,30). Further, psychological assessments are also carried out to assess the improvement in their psychological status and psychological influences of pain. However, none of the instruments in assessing pain, function, or psychological status provide a reliable assessment.

Academicians continue to discuss the role of pain assessment and change in the functional status and interrelationships between pain and function. Overall, function is considered as a stronger reliable indicator. In fact, similar to pain, assessments of function can be dependent on a multitude of factors. To avoid various conflicts, automatic pain detection has been promoted in recent years, which has gained some traction since the passage of the American Recovery and Reinvestment Act of 2019, which has mandated the electronic medical record (EMR)/electronic health record (EHR), and resulted in explosive growth. However, EMRs have been associated with numerous disadvantages, including increasing costs, loss of patient contact, irrelevant information, increase in litigation, fraud and abuse implications due to the common practice of copy, paste, and auto-populating (31-36).

Artificial Intelligence and Machine Learning

AI is the ability of machines to exhibit human-level intelligence and mimic human actions. Although the evolution of AI can be traced back to 1956, the concept really began to flourish in the last decade due to a huge boost in performance and speed with the help of low-cost high-performance graphics processing units (GPUs), cheaper storage options, faster cloud computing, parallel computing advancements coupled with a surge in data collection (37-40). The initial definition of AI originating from Turing, who proposed an experiment where 2 players, who can either be human or artificial, try to convince a human third player that they are also humans (41). Thus, the test of AI is considered to be successful if the third player cannot tell who the machine is. Important steps in the development of machine learning (ML) for the first creation of the computer learning program, which was a checker game (42), and the first neural network called the perception (43). Thus, AI is understood as a field of study that combines computer science, engineering, and related disciplines to build machines capable of behavior that would be said to require intelligence were it to be observed in humans (39). Such behaviors may be described as the ability to visually perceive images, recognize speech, translate language, and learn from and adapt to new information (39). To achieve these goals, AI as a field of study, can employ a number of techniques, including ML. As an example, ML allows algorithms to make predictions and solve problems based on large amounts of data, without being explicitly programmed (39). In addition, another subset of ML is deep learning, which goes further to using multiple layers of artificial neural networks to solve complex problems from unstructured data, much like the human brain (39,44). Thus, AI has been described as the fourth industrial revolution with transformative and global implications (38).

Artificial Intelligence

AI is described in 3 categories. The first one is artificial narrow intelligence (ANI), also known as weak AI. It is the ability of a machine to specialize and solve a specific problem. These systems are goal oriented and can often perform on par or better than humans in specific problems under certain conditions. This is the current available form of AI used in voice assistant systems in smartphones such as Apple's Siri and Amazon's Alexa. The second type of AI is artificial general intelligence (AGI), also known as strong AI. It is the ability of a machine to reach general human cognitive function across a wide variety of domains. The machine can learn and solve a variety of problems and can act in a way that is indistinguishable from a human. Lastly, the third type, artificial super intelligence (ASI), is a hypothetical state where the machine surpasses all human capabilities and becomes self-aware. This is the successor to AGI. We have not reached AGI or ASI yet.

Machine Learning

ML, natural language processing (NLP) and deep learning (DL) are subsets of AI. ML is the science where computers learn from given data without being explicitly instructed/programmed. ML can be described as an intersection between statistics and computer science. A ML algorithm learns by iterating over a dataset, recognizing patterns and making predictions on new data. DL is a subset of ML; inspiration for the design and structure of the algorithm is derived from the human brain. DL algorithms have gained popularity in the last decade due to their state-of-the-art performance in tasks such as image classification on the ImageNet competition. NLP is the ability of a machine to understand text and speech like a human being. It is a combination of rule based computational linguistics and ML methods. NLP has found applications in tasks such as speech recognition, sentiment analysis, voice assistants, chatbots, language translation, etc.

There are three major types of learning mechanisms in ML. The first one is supervised learning (SL) which uses labeled data along with the inputs to train the algorithm to learn to classify or predict the value for the label (output). SL has two methods: classification and regression. The input data is used to tune the weights iteratively until the model is fitted according to the requirements. The second type is unsupervised learning (UL) where the algorithm discovers data grouping and hidden patterns in the data without human intervention. The third type is reinforcement learning (RL) where an intelligent agent learns actions to maximize reward in a situation with constraints and parameters. The agent learns to find an optimal path through a series of rewards and punishments after each attempt.

ML has gained general interest and has penetrated daily life in recent years. Healthcare data is being collected in wearable devices and smartphones at a rapid pace (45-49). Devices like the Apple watch can monitor sleep through the detection of movements and the identification of cardiovascular irregularities such as arrhythmia, which may suggest atrial fibrillation (45).

ML has been used in multiple medical specialties including pain management, neurosurgery, orthopedic surgery, anesthesiology, radiology, oncology, neurology, psychology, and psychiatry among others (50-64). While interventional pain management is defined as, "the discipline of medicine devoted to the diagnosis and treatment of pain related disorders principally with the application of interventional techniques in managing sub-acute, chronic, persistent, and intractable pain, independently or in conjunction with other modalities of treatment" (65), it incorporates multiple specialties, including anesthesiology, neurosurgery, radiology, and orthopedic surgery as patients are treated by multiple specialists. In fact, many of the traditionally available tests can be utilized with AI.

Multiple outcome assessments have been utilized extensively in surgical literature, as well as interventional pain management (1,50-64).

Lötsch and Ultsch (1) have performed an extensive review of the available literature of ML in pain research published in 2018. At the time, they found 88 original reports of the use of ML in a pair context. Overall, 52 reports met inclusion criteria for several different methods of ML in pain research. They identified and described pain phenotype prediction from complex case data, structure detection in complex pain-related data, knowledge discovery and exploration of pain-related data.

Rashidi et al (56) described ML with utilization of large data set analysis to individualize pain management. They reviewed core principles and definitions in the field of ML. They also examined the impact of ML approaches in the analyses of large EHR data sets. Subsequently, the authors reviewed advanced machine and deep learning approaches to semi-structured and unstructured datasets, highlighting future directions in the use of ML to enhance our understanding and treatment options for patients with pain. Rashidi et al (56) extensively defined and describes tasks of ML tasks, measurements of ML performance, early explorations, and ongoing challenges and opportunities. They cautioned that despite the myriad advances offered by ML, the new analytical techniques have also forced reckoning by physicians and researchers on fundamental challenges concerning the application of evidence-based medicine to individual patients reckoning. Consequently, it is crucial that physicians must remember that patients are more than just data. Accurate ML learning may enhance disease diagnosis, but they cannot deliver that diagnosis with compassion and understanding, with the recognition of the impact of that diagnosis on the patient and their future. They also advised that one of the latest potential benefits of ML in medicine would be allowing physicians to attend more to the humanistic needs of our patients.

Methods

The role of ML in pain research has been extensivelyreported. Lötsch and Ultsch (1) described that the emerging discipline of computational pain research provides contemporary tools to understand pain. Computational pain research uses computer-based processing of complex pain-related data and relies on intelligent learning algorithms. Consequently, in this review, we review the current AI and ML technology standards at practice in health care and to disseminate knowledge and potential benefits and future applications, specifically in pain management. Figure 1 shows the Venn diagram showing the relationship between AI, ML, DL, and NLP. The methodology included a narrative review with a comprehensive review of the literature pertinent to pain management, including pain assessment.

Objectives

The objectives of this review are to collect, analyze, and summarize available peer-reviewed studies, which use ML algorithms to find gaps in research and pave the way for future research.

Search Strategy

Research papers were searched using a literature search on PubMed, Science Direct and IEEE Xplore between 2018 and 2020. The keywords used to search the articles were: "pain assessment", "outcomes assessment", "functional status assessment", "pain prediction", "pain intensity estimation", "machine learning", "deep learning", "facial pain", "pain recognition" and "pain management" to be present in title or abstract.

Inclusion Criteria

Inclusion criteria are as follows:

- The paper is written in English
- The paper includes AI or ML methods
- The paper addresses any stage in pain care management
- Only studies that have been performed on humans were included

Exclusion Criteria

Exclusion criteria are as follows:

- Review papers were excluded
- Papers with content overlap (same dataset and methods) were excluded
- Papers that do not include performance metrics of the algorithms used were excluded
- Chest pain, cancer pain, migraine and headache studies were excluded

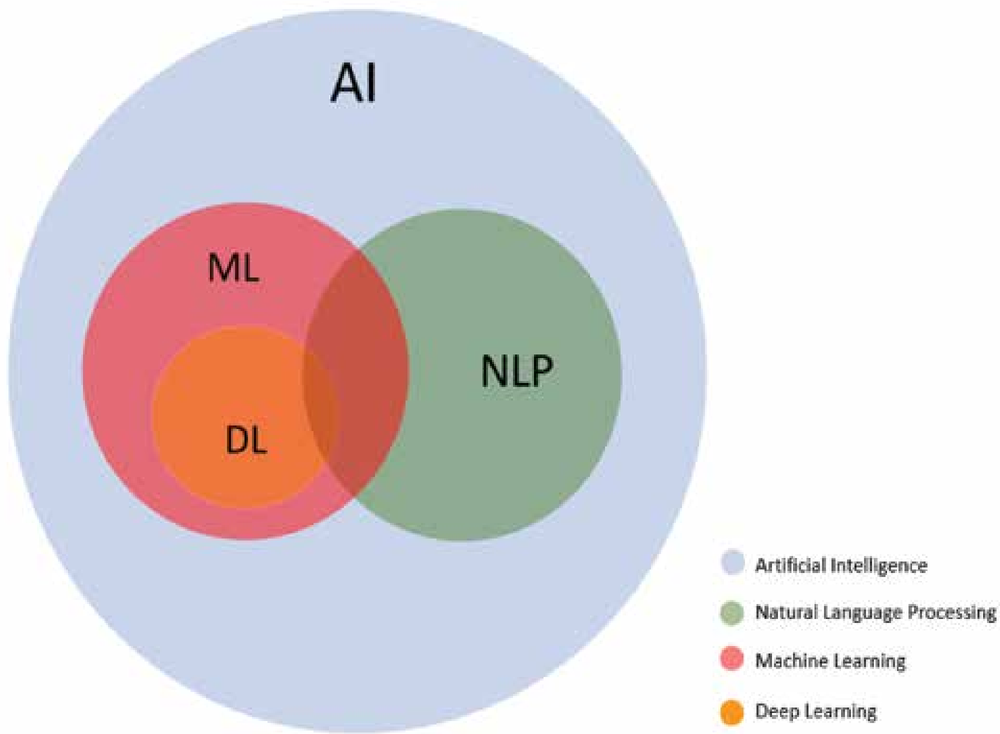

Fig. 1. *Venn diagram showing relationship between AI, ML, DL and NLP.*

Results

The results of the literature search are shown in Fig. 2. After review, of 350 studies, 57 were considered for inclusion. Of these, 47 studies met inclusion criteria (57,59,61,66-109).

Overall, 47 studies (57,59,61,66-109) were included. Of these, 11 studies were for pain assessments (57,68,70,71,75,90,100-102,104,105), 7 studies for facial expression (66,67,69,71,82,89,103), 2 studies for neonatal pain assessment (74,99), 5 studies for pain assessment in special settings (73,76,79,93,95), 10 studies for automated measurements (78,80,81,83-86,88,94,96), 8 studies for spinal diagnosis (77,87,92,97,106-109), and 4 studies for treatment algorithms (59,61,91,98).

Pain Assessments

Multiple investigators evaluated pain utilizing various methods.

Anan et al (57), in a small randomized controlled short-term study, evaluated the effects of an AI assisted health program in workers with neck and shoulder pain/stiffness and low back pain. Although small, this trial study's aim was to evaluate improvements in musculoskeletal symptoms in workers with low back pain and neck/shoulder stiffness/pain following their use of an exercise-based AI-assisted that operated through a mobile messaging app. This study included the analysis of 48 patients in the intervention group and 45 in the control group with an adherence rate of 92% (44/48) in the intervention group who showed significant improvements in the severity of neck/shoulder pain/stiffness and low back pain in comparison to the control group (OR 6.36, 95% CI 2.57-15.73, $P < .001$). Utilizing subjective assessment of improvement in pain/stiffness at 12 weeks, 36 (75%) of 48 patients in the intervention group and 3 (7%) of 46 in the control group reported improvements (either improved or slightly improved) (OR 43.00, 95% CI 11.25-164.28, $P<.001$). Although small, this study shows that the short exercises prompted by the AI-assisted health program improved both neck/shoulder pain/stiffness and low back pain at 12 weeks follow-up. Further, multiple studies are needed to replicate these findings and identify the contributing elements to the successful outcome of the AI-assisted health program.

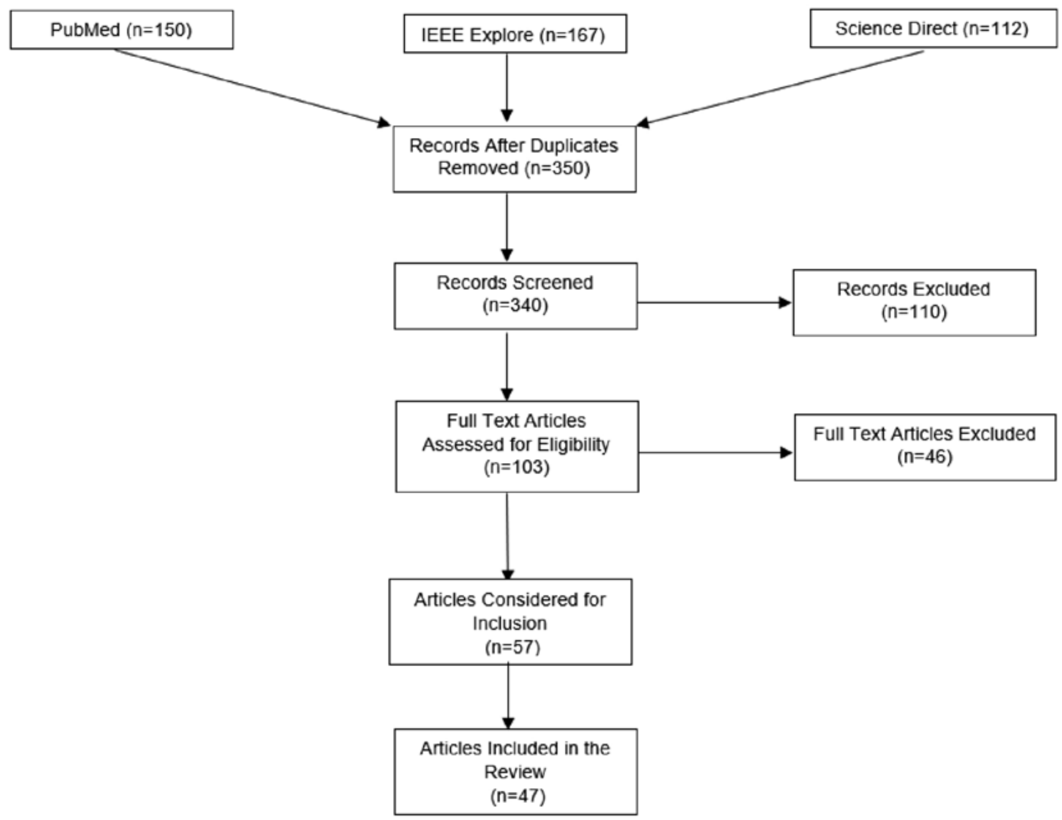

Fig. 2. *Flow diagram of the literature review.*

Zhao et al (68) evaluated a deep learning framework for chronic pain score assessment. This study used a plethora of sensors on patients to assess chronic pain. The pain was measured in 2 datasets, the 1st dataset from a chronic pain patient who had been suffering for 10 years and the 2nd dataset was from a group of chronic pain patients. Dataset 1 had 2 classes while dataset 2 had 7 classes. Sensors used were photoplethysmography (PPG), skin temperature, galvanic skin response (GSR), accelerometer and gyroscope. Algorithms used were Ordinal Regression, Convolutional Neural Network (CNN), Multilayer Perceptron and Logistic Regression (LR). The results showed that it is possible to objectively classify chronic pain using ML methods with CNN performing the best with an accuracy of 95.23%. With this complicated technology and a time-consuming process, the authors concluded that this is a proof of principle for chronic pain score assessment via deep learning. They also believed that it can provide an objective pain assessment for each patient. However, the study is hampered by a small number of patients and only one patient in one group and the complexity needed for the testing.

Susam et al (70) examined automated pain assessment using electrodermal activity (EDA) data and ML. They postulated that automated pain detection from physiological data may provide important objective information to better standardize pain assessment. They also theorized, specifically, that EDA can identify features of stress and anxiety induced by varying pain levels. Consequently, in this study, they used time scale decomposition (TSD) to extract the salient features from EDA signals in order to identify an accurate and automated EDA pain detection algorithm which could sensitively and specifically distinguish pain from non-pain conditions. Similar to multiple other studies, this is a very small study with 21 neurotypical youth, 16 men and 5 women, primarily Hispanic (71%) and with a median age of 11 years were included while they were undergoing laparoscopic appendectomy. TSD was used to extract features from the EDA data followed by classification into pain (moderate-to-severe pain)/no pain classes using the Support Vector Machine (SVM) algorithm. The results showed that EDA is a useful metric for pain assessment and has achieved an accuracy of 77.66% in recognition of pain compared to no pain classes. While the authors felt that this is the first study to apply TSD to EDA data in this analysis, they also believed that they completed TSD using scale decomposition (SD) on each high-quality, normalized, filtered, down sample EDA signal. Based on the accuracy score of 77.66% in an extremely small study of 21 patients, the authors felt that this paper contributed to the utility of TSD as a novel feature extension method for use with EDA data and the results also represented promising preliminary evidence for an accurate ML classification algorithm to discriminate clinically moderate to severe pain versus no pain in children using EDA patterns alone. The limitations as described include not only a smaller number of patients included, but also the complex nature of the measurements. This technology may be of use in future developments with large-scale studies.

Fodeh et al (72) assessed the value of ML with classification of clinical notes with pain assessment. The dataset included patients with documented pain intensity ratings of above 4 and initial musculoskeletal diagnosis captured by the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes in fiscal year 2011 and a minimal one year of follow-up lasting as much as 3 years maximum. A total of 92 patients with 1,058 notes were utilized. Based on their schematic documentation, they found variations in documenting the subclasses of pain assessment. The variations were observed in pain site (67%), intensity of pain (57%), persistence of pain (32%), etiology of pain (27%), and documentation of patients' reports of factors that aggravate pain was only present in 11% of positive notes. The authors developed a random forest classifier to identify clinical notes with pain assessment information. The random forest classifier achieved the best performance labeling clinical notes with pain assessment information, compared to other classifiers: 94%, 95%, 94%, 94% in terms of accuracy, positive predictive value, F1-score and Area Under the Curve (AUC), respectively. The major advantages of this study are the study was performed without any additional effort and 92 patients with a moderate sample size were included with a long-term follow-up. However, lack of appropriate data in many patients is of concern in the documentation patterns.

Santana et al (90) studied 338 controls (no pain) and 659 chronic pain patients (440 fibromyalgia and 219 chronic back pain patients). Five datasets were created using questionnaires related to depression, anxiety, heat and cold threshold, pressure stimulus, etc. The paper focused on identifying chronic pain syndromes using different ML algorithms. Models used to classify are: LR, SVM, K-Nearest Neighbors (K-NN), Dynamic tree, Random Forest, Extra trees classifier, Multi-layer Perceptron, XGBoost and Neural networks. The paper concludes that ML algorithms performance correlates with larger data size, hyper-parameter tuning, and type of ML model used. Of the available algorithms, the ensemble models performed better after fine-tuning. Extra Trees Classifier performed the best with a score of 0.793 for mean balanced accuracy, best score for mean Area Under the Receiver Operator Curve (AUROC) was 0.876 using the XGBoost model. This study in larger patient population described with ensemble models shows promise.

Haque et al (75) described a new pain dataset, Multimodal Intensity Pain (MIntPain) and two methods of using the data to classify the dataset into pain classes in 20 healthy volunteers subjected to electrical impulses. There are cameras for color video red, green, blue (RGB), Thermal and Depth. It is also referred to as a red, green, blue, depth and thermal (RGBDT) dataset. There are 5 classes, 0 for 'no pain' and 1 to 4 for the pain classes. The data is used for pain classification using a standalone CNN algorithm and a hybrid CNN+LSTM algorithm. The CNN algorithm looks at spatial features in video frames while the CNN+LSTM hybrid looks at Spatial and Temporal features associated with the video. There are two methods for the hybrid algorithm: Early fusion and Late fusion. Early fusion is when the RGB, depth and thermal video frames are all stacked together to create a 5-channel image which is then passed to the CNN algorithm and the output features from the CNN are then sent to the LSTM model. Late fusion is where there are 3 separate CNNs, each processing RGB, depth and thermal separately and then all the acquired features are sent to the 3 separate LSTMs and finally the output is fused, and a early fusion is used for classification. Baseline results were provided, and the best results were achieved by the early fusion RGBDT method with 31.40% for per frame accuracy and 36.55% for per sequence accuracy. The study described a complicated design in 20 healthy volunteers the may not be applicable in clinical settings.

Chesler et al (100) retrospective identified and ranked sources of variability in nociceptive responses occurring over a several year period in a typical research laboratory using a computational approach. Out of their archival data set of 8,034 independent observations of baseline thermal nociceptive sensitivity, they applied a machine-learning algorithm. What was revealed in this analysis was that there was a factor even more important than mouse genotype was the experimenter actually performing the test. They further found that nociception can be affected by many additional laboratory conditions including: season/humidity, cage density, time of day, sex, and within cage order of testing. Their results were confirmed using linear-modeling in a subset of their data, as well as in confirmatory experiments in which they were able to partition the variance of this complex to genetic (27%) environmental (42%), and genetic X environmental (18%) sources. This study is based on evidence that, in biobehavioral experiments, laboratory conditions are commonly assumed to be "controlled" with little impact on the outcome. However, recent studies have illustrated that the laboratory environment has a robust effect on behavioral traits. Also, it is widely accepted that environmental factors can interact with trait-relevant genes. Furthermore, the generalizability and reliability of behavior genetic research which has been designed to identify those genes is under debate. Chesler et al (100) in this review utilizing a large archival base observed genetic as well as environmental sources contributing to the variance by linear modeling. It appears to be a straightforward study and may have significant clinical applicability.

Huang et al (101) were looking to identify an optimal set of characteristics for supporting self-management. To do so, they used a ML approach to analyze self-reporting data which had been collected from an integrated biopsychosocial treatment program. Additionally, they proposed a classification model to differentiate stages of treatment. They applied 4 different feature selection methods to rank the questions, and utilized 4 supervised learning classifiers to investigate the relationship between the numbers of questions and classification performance. The results showed no significant difference between the feature ranking methods for each classifier in the overall classification accuracy or AUC ($P > 0.05$). However, there were significant differences between the classifiers for each ranking method ($P < 0.001$). Overall, the results showed that the multilayer perception classifier had the best classification performance on an optimized subset of questions consisting of 10 questions. Its overall classification accuracy and AUC were 100% and 1 respectively. Thus, this study shows 100% accuracy with simplified questionnaires and has clinical applicability for the near future.

Bui and Zeng-Treitler (102) sought in this study to automate both the creation and utilization of regular expressions in text classification. To accomplish this, they designed a novel regular expression discovery (RED) algorithm and subsequently implemented two text classifiers based on RED. They used two clinical datasets for testing and evaluation. The first was the SMOKE dataset, with 1,091 text snippets which describing smoking status, and the second was the PAIN dataset with 702 snippets describing pain status. They then performed 10-fold cross-validation to determine accuracy, precision, recall, and F-measure metrics. As part of the evaluation, SVM classifier was trained as the control. The results of this study were that in overall accuracy on the 2 datasets, the 2 RED classifiers achieved 80.9-83% which was 1.3-3% higher than the accuracy of the SVM. Also, small but consistent improvements were noted in precision, recall, and F-measure when comparing RED classifiers to SVM alone. More importantly, RED + ALIGN classified correctly many cases that were misclassified by the SVM (8.1-10.3%) of total instances and 43.8-53.0% of SVMs misclassification. They concluded that machine generated regular expressions can be effectively used for clinical text classification. Additionally, classification performance can be improved by combing the regular expression based classifier with other classifiers such as SVM. As the authors goals was to automate both the creation and utilization of regular expressions in text classification, it appears that they have been successful with multiple text snippets with different types of classifications. Further, study with replication of the data and methodology may be applicable to clinical settings.

Lötsch et al (104) studied the pattern of neuropathic pain induced by the topical application of capsaicin in healthy volunteers in an attempt to better reflect clinical pain conditions. In this attempt they used the Quantitative Sensory Testing (QST) battery on untreated ("control") and topical capsaicin-hypersensitized ("test") skin. They then compared Z-transformed QST-parameter values they obtained with corresponding values published from 1,236 patients with neuropathic pain using Bayesian statistics. Patients were then clustered for the resemblance of their QST pattern to neuropathic pain. In spite of the fact that the QST parameter values from the untreated site agreed with reference values, there were several QST parameter of those treated with topical capsaicin deviated from normal, resembling in 0 to 7 parameters of the QST patterns in patients with neuropathic pain. In 18% of the patients, degrees of resemblance as high as 50-60% to neuropathic pain were obtained utilizing a classification and regression tree composed of 3 QST parameters (mechanical pain sensitivity, wind-up ratio, and z-transformed thermal sensory limen), inclusion in the respective clusters were predictable at a cross-validated accuracy of 86.9%. They concluded that topical capsaicin partially produced the desired clinical similarity to neuropathic pain in a pre-selectable subgroup of healthy patients to such as degree as to encourage the expectation that experimental pain models can be optimized toward mimicking clinical pain. These patients therefore qualify for enrollment in analgesic studies utilizing highly selected cohorts to enhance predictivity for clinical analgesia. The authors in this study base it on the fact that human experimental pain models are widely used to study drug effects under controlled conditions. However, these conditions and available literature require further optimization to better reflect clinical pain conditions. It should be noted that this study included 110 healthy volunteers and the best results were 50-60^ of resemblance to neuropathic QST patterns obtained in only 18% of the patients. Consequently, while this study may have potential clinical applications in the future, further improvements in the methodology may show relevant clinical applications in the future.

Dimova et al (105) in this observational study, the hypothesis that an inducible subgroup would differ from health patients with regard to their psychological phenotype was pursued. The patients were assessed using a comprehensive set of variables compared of general psychological and pain-related cognitive-emotional mechanisms. Using the sum scores from the questionnaires, a significantly linearly correlation with each other was found. The major source of variance indicated by principal component analysis was 46% and was attributed to dispositional optimism measured by the Life Orientation Test (LOT) which also significantly differed between the groups, either those in whom a neuropathy-like pattern of pain as assessed by the QST could be partially (50-60% of the 11 QST parameters) induced (n=20) or not (n=90; *P*=.0375). This dispositional optimism appeared again as the amin selection factor in a classification and regression tree predicting a patients group assignment (inducible neuropathy – like QST pattern versus non-inducible neuropathy-like QST pattern) with a cross-validated accuracy of 95.5 ± 2.1%. The conclusion was that the few patients in this random sample of healthy volunteers who, after the application of topical capsaicin partially resemble (about 60%) the clinical pattern of neuropathic pain in the QST battery, and are pre-selectable on the basis of psychological factors, particularly that of an emphasis on pessimistic life attitudes. Dimova et al (105) hypothesized that clinical patterns of neuropathic pain, diagnosed using the QST batter, could be partly produced in healthy volunteers by the typical application of capsaicin. They further hypothesized that this type of induced pain may be similar to the neuropathic pain that develops in a sub-group of patients who have neurologic lesions, but which showed a correlation only in a small fraction of healthy individuals. In this study they showed only a partial success rate. However, this continues to be an experimental design in healthy volunteers and consequently any results are not applicable into clinical settings until they are replicated and reproduced on multiple occasions.

Table 1 summarizes literature of pain research with application of machine learning and/or artificial intelligence for pain assessments.

Facial Expression

In the field of pain management, facial pain can be measured using metrics like the PSPI score, Action Units, etc., The Facial Action Coding System is a system used for understanding facial movements, broken down into individual metrics of muscular changes called Action Units (67). PSPI score is a measure of pain as a combination of Action Unit intensities and detection score (69). Manual coding of Facial Action Coding System (FACS) is performed by FACS certified professionals who undergo many hours of training. The human coder looks at every frame in the video to find the presence and then to measure the intensity of the Action Units along with their onset and offset markers. Videos are usually shot between 24 and 60 frames per second and manually annotating every frame even in a short video recording is a time-consuming and expensive process (69). Automatic methods using ML methods for facial action unit coding are being developed to solve the problem of manual coding. OpenFace is a tool for automatic detection and estimation of action units in real-time to be used for pain analysis (71).

Prkachin and Solomon (67) in 2008 studied the structure, reliability, and validity of facial expressions

Table 1. *Reports of pain research with application of machine learning and/or artificial intelligence for pain assessments.*

Author, Year **Title** **Journal/Source**	**Data Analysis And Machine Learning Methods**	**Conclusion/Comments**
Anan et al (57), 2021 Effects of an artificial intelligence-assisted health program on workers with neck/shoulder pain/stiffness and low back pain: randomized controlled trial. JMIR Mhealth Uhealth 2021; 9:e27535.	Exercise based artificial intelligence	This small study showed that the short exercises provided by the AI assisted health programs improved both neck and shoulder pain stiffness and low back pain at 12-week follow-up.
Zhao et al (68), 2020 How much does it hurt: a deep learning framework for chronic pain score assessment. 2020 International Conference on Data Mining Workshops (ICDMW). Sorrento, Italy, 2020, pp 651-660.	Deep learning framework	Authors concluded that this is a proof of principle for chronic pain score assessment via deep learning. This study is hampered by a small number of patients and only one patient in one group and the complex need for testing.
Susam et al (70), 2018 Automated pain assessment using electrodermal activity data and machine learning. Annu Int Conf IEEE Eng Med Biol Soc 2018; 2018:372-375.	Machine learning classifier	Authors concluded that electrodermal activity data can be used independently to discriminate between No pain vs pain (Moderate and severe) with high accuracy. This study also explores a novel feature extraction method for EDA data analysis.
Fodeh et al (72), 2018 Classifying clinical notes with pain assessment using machine learning. Med Biol Eng Comput 2018; 56:1285-1292.	Machine learning classifier	This paper showed an automated pain assessment method using machine learning on unstructured clinical notes from EMR with good accuracy. This method can detect the presence of pain from the notes but does not give qualitative or quantitative information of the detected pain experience.
Santana et al (90), 2020 Chronic pain diagnosis using machine learning, questionnaires, and QST: A sensitivity experiment. Diagnostics (Basel) 2020; 10:958.	Machine learning and deep learning classifiers	For chronic pain classification, the researchers concluded that ensemble ML methods, greater processed data, and higher optimization presents a higher chance of diagnostic success. A limitation from this study included a shortage of data sources.
Haque et al (75), 2018 Deep multimodal pain recognition: A database and comparison of spatio-temporal visual modalities. 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018) Xi'an, China, 2018, pp 250-257.	Deep learning architecture	This paper presents a multimodal dataset for pain assessment with color, depth, and thermal video data. The authors have also developed a pain recognition model using various data fusing approaches and conclude that multiple visual data are useful for pain recognition.
Chesler et al (100), 2002 Identification and ranking of genetic and laboratory environment factors influencing a behavioral trait, thermal nociception, via computational analysis of a large data archive. Neurosci Biobehav Rev 2002; 26:907-923	Machine Learning Algorithm	A machine learning algorithm was applied to a dataset consisting of 8034 observations of pain sensitivities in mice. The factors that affected the outcomes of the experiments were ranked in terms of genetic and laboratory environmental influence.
Huang et al (101), 2011 Feature selection and classification in supporting report-based self-management for people with chronic pain. IEEE Trans Inf Technol Biomed 2011; 15:54-61.	Machine Learning Algorithm	187 patients with chronic pain took a lengthy questionnaire whose results were plugged into a ML algorithm to isolate the most influential questions for diagnosis. This optimizes questions most useful for a self-management system.

Table 1 (cont.). *Reports of pain research with application of machine learning and/or artificial intelligence for pain assessments.*

Author, Year Title Journal/Source	Data Analysis And Machine Learning Methods	Conclusion/Comments
Bui & Zeng-Treitler (102), 2014 Learning regular expressions for clinical text classification. J Am Med Inform Assoc 2014; 21:850-857.	Natural language processing and machine learning algorithms	This paper shows a method to automate the creation and utilization of 'regular expressions' for clinical text classification using natural language processing techniques and the methods have shown improvement over a popular machine learning algorithm.
Lötsch et al (104), 2015 Pattern of neuropathic pain induced by topical capsaicin application in healthy patients. Pain 2015; 156:405-414.	Machine Learning Algorithm	The study looks at 110 healthy volunteers who were subjected to experimentally induced pain and concluded that topical capsaicin stimulated a pattern of neuropathic pain. This displays the utility of experimental pain models in imitating clinical pain.
Dimova et al (105), 2015 A more pessimistic life orientation is associated with experimental inducibility of a neuropathy-like pain pattern in healthy individuals. J Pain 2015; 16:791-800.	Statistical Analysis	The authors studied 110 individuals with differing psychological compositions who were subject to capsaicin stimulus. Those who harnessed a more pessimistic life attitude were more likely to exhibit a neuropathic pain pattern.

of pain. They used a sample of 129 patients with shoulder pain to whom they administered a series of active and passive range-of-motion tests to both affected and unaffected limbs which were then repeated at a second time. Then, using three self-reported scales, the patients rated the maximum pain induced by each test. The Facial Action Coding System was used to simultaneously measure facial actions and found that several facial actions were able to discriminate painful from non-painful movements. Based on these actions, an index of pain expression was able to demonstrate test-retest reliability as well as concurrent validity with self-reports of pain. Their findings support the idea of core pain expression that has desirable properties as well as being consistent with the suggestion of individual differences in pain expressiveness. They also provided a discussion of reasons for variance in previous studies between pain expression and self-reports.

Bargshady et al (82) studied pain intensity detection using facial expressions with the help of an Ensemble Deep Learning Model. The datasets used were Mint Pain and UNBC-McMaster. The algorithm fuses 3 different streams of data into the model. The pain output consisted of 5 labels to denote the pain level intensity. RGB frames, Depth frames as well thermal image data are fed into a CNN-recurrent neural network (RNN) architecture for spatial-temporal analysis. The CNN used here is a pre-trained VGG Face CNN and the RNN used here is an LSTM. Ensemble of learning algorithms are composed of multiple weak learners which together will predict with higher accuracy. A weak learner is a learner that performs better than random guessing. In the early fusion, CNN and PCA are working together to extract and select features. These features are then transferred into late fusion for classification. The model contains 3 independent CNN-RNN deep learners. The achieved state of the art accuracy and performance metrics. The results were compared to a baseline VGGFace + 1 stream LSTM. The accuracy of multi-level pain detection is around 89% and the AUROC is 0.93.

Bargshady et al (89) sought to develop an enhanced deep learning algorithm to detect pain intensity from facial expression images. In this study they reported on the development of a new enhanced deep neural network which was designed to be more effective framework for detecting pain intensity. To accomplish this, they used a 4 level threshold using images of facial expression. They utilized the UNBC-McMaster Shoulder Pain Archive Database of facial images which was first balanced and then used for the training and testing of the classification system; which was paired with the VGG-face pre-trainer as the featured tool for future extraction. In order to improve the computational efficiency and reduce the dimensionality of their classification model and extract the most relevant features, they used Principle Component Analysis. The pre-screened features which had been used as model inputs, were then transferred to produce a joint hybrid (EJH)-CNN-BiLSTM) deep learning algorithm composed of CNN which were then linked to the joint hybrid by LSTM for multiclassifcation of pain. Overall, the results showed that the EJH-CNN-BiLSTM model tested to estimate 4 different levels of pain and revealed good accuracy in terms of different performance evaluation technique. The results also indicated that the enhanced EJH-CNN-BiLSTM classification algorithm as explored showed potential to be used an an AI tool for the medical diagnosis in automatic detection and therefore usefulness for the subsequent management of patients.

Overall, Bargshady et al (82,89) describe that the automated detection of pain intensity from facial expression, especially from facial images that show a patient's health, remains a significant challenge in the medical diagnosis and health informatics area. Expert systems that prudently analyze facial expression images, utilizing an automated ML algorithm, can be a promising approach for pain intensity analysis in the health domain. The 2 studies described above add to the significant research within the pain recognition and management area that aim to adapt facial expression datasets into deep learning algorithms to detect pain intensity in binary classes, and also to identify pain and non-pain faces. Overall, facial expression tools may be clinically applicable in the future with refinement and replication of the data.

Meng et al (103) in their study sought to improve the recognition performance of continuous naturalistic affective expression using an dinformation theory approach. They used the datasets of naturalistic affective expression (AVEC 2011 audio and video datasets, PAINFUL, video dataset) continuously labeled over time as well as different dimensions to analyze the transactions between levels of those dimensions to show that these transitions occur very slowly and because of that they recommend modeling them as first order Markov models. With regard to the dimension levels they are considered to be the hidden states in the Hidden Markov Model (HMM). Using the labels provided with the training set, their discrete transition and emission matrices are trained. Then, the recognition problem is converted into a best path-finding problem in order to obtain the best hidden states sequence in HMMs and this is a key difference from the previous use of HMMs as classifiers. Next, there is integration of the transitions between dimension levels using a multistage approach in which they first level performs a mapping between the affective expression features and a soft decision value and further classification stages are modeled as HMMS that refine that mapping by taking into account the temporal relationship between the output decision labels. By taking into account the temporal relationship, the experimental results for each of the unimodal datasets show overall performance to be above that of a standard classification system that do not do so. Specifically, the results of the AVEC 2011 audio dataset outperformed all other systems presented at the international competition. Meng et al (103) based their paper on the available literature that naturalistic affective expressions changed at a rate much slower than the typical rate at which video or audio is recorded. They hypothesized that this phenomenon increases the probability that consecutive recorded instances of expressions represents the same affective content. They explored these issues in their report. Their results and conclusions are valid, raising further questions on facial expression technology to apply in clinical pain management without further analysis and refinement of the technology.

In contrast to the above reports describing the reliability of facial expression, Bartlett et al (66) reported on deceptive pain expressions and their decoding. This is a development whereby even with AE and ML technology, deceptive pain expressions can be identified. The human face has evolved to convey rich information for social interaction which includes the expression of both emotions and pain (66). Facial movements are controlled by 2 motor pathways; a subcortical extrapyramidal motor system which drive spontaneous facial expression of felt emotion, and a cortical pyramidal motor system that controls voluntary expression. These voluntary deceptive expressions are so successful that Bartlett et al (66) reported that they can deceive most human observers, whereas, by identifying subtle differences between pyramidally and extrapyramidally driven movements machine vision may be able to differentiate deceptive from genuine facial signals. Further, human observers could not differentiate real from fake expression of pain better than by chance and even after training only improved to a modest 55% accuracy. On the other hand, the computer vision system by automatically measuring facial movements and performing pattern recognition attained 85% accuracy. The difference was attributed to the machine systems superior ability to differentiate the minute dynamics of genuine and fake expressions that were unobservable to the human observers.

These findings of facial expression technology are encouraging and with appropriate replication of the data and additional clinical studies, facial expressions may be a technology useful for clinical applications.

Table 2 summarizes the literature of pain research with the application of machine learning and/or artificial intelligence for patient facial expression.

Neonatal Pain Assessment

Pain assessment in patients who are unable to verbally communicate is a challenging problem (99). The fundamental limitations in pain assessment in neonates stems from subjective assessment criteria, rather than quantifiable and measurable data (99). This often results in poor quality and inconsistent treatment of

Table 2. ***Reports of pain research with application of machine learning and/or artificial intelligence for patient facial expression.***

Author/Year	**Data Analysis And Machine Learning Methods**	**Conclusion/Comments**
Prkachin & Solomon (67), 2008 The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain. Pain 2008; 139:267-274.	Statistical analysis	This study observes the properties of facial expressions of 129 people for pain analysis. The authors concluded that a specific set of facial expressions contribute to the core expression of pain.
Bargshady et al (82), 2020 Ensemble neural network approach detecting pain intensity from facial expressions. Artif Intell Med 2020; 109:101954.	Deep learning algorithm	The authors describe a novel ensemble algorithm which uses visual facial features to identify multi class pain levels (5 levels). This algorithm uses a combination of convolutional neural networks as well as sequential models to classify the pain levels.
Bargshady et al (89), 2020 Enhanced deep learning algorithm development to detect pain intensity from facial expression images. Expert Syst Appl 2020; 149:113305.	Deep learning algorithm	The authors describe a deep learning model to identify 4 pain intensities from facial expressions using a hybrid bidirectional algorithm. The limitation described in the paper is the lack of standardized datasets for facial pain detection.
Meng et al (103), 2014 Affective State Level Recognition in Naturalistic Facial and Vocal Expressions. IEEE Trans Cybern 2014; 44:315-328.	Statistical analysis	The authors describe a method to recognize the pain intensity levels by understanding the relation between levels of affective dimensions using a statistical model on audio and video data.
Bartlett et al (66), 2014 Automatic decoding of facial movements reveals deceptive pain expressions. Curr Biol 2014; 24:738-743.	Machine learning algorithm	This paper demonstrates an automatic computer vision algorithm which can detect faked pain expression from video data (by understanding the subtle differences between expressions). The algorithm outperformed human observers by a significant margin.
Walecki et al (69), 2016 A framework for joint estimation and guided annotation of facial action unit intensity. 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) Las Vegas, NV, 2016, pp 1460-1468.	Machine Learning	The authors report a coding framework that can assist manual facial coding by focusing in on the most significant predictors. A faster classification process allows for more efficient annotations in fields such as facial pain recognition.
Baltrusaitis et al (71), 2018 OpenFace 2.0: Facial behavior analysis toolkit. 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018) 2018, pp 59-66.	Deep Learning Algorithm	This study describes a real-time advanced software which can accurately identify a greater scope of facial characteristics than previous tools. This tool bridges the use of facial muscle activations for pain assessment purposes.

patient pain management. Recent advancements in pattern recognition techniques using relevance vector machine (RVM) learning techniques can assist medical staff in assessing pain by constantly monitoring the patient and providing the clinician with quantifiable data for pain management. The RVM classification technique is a Bayesian extension of the SVM algorithm, which achieves comparable performance to SVM while providing posterior probabilities for class memberships and a sparser model. If classes represent "pure" facial expressions (i.e., extreme expressions that an observer can identify with a high degree of confidence), then the posterior probability of the membership of some intermediate facial expression to a class can provide an estimate of the intensity of such an expression (99).

Gholami et al (99) used relevance vector machine (RVM) classification technique to determine pain from non-pain in neonates as well as to assess the level of pain intensity, while at the same time correlating the results with the pain intensity assessment of expert and nonexpert human examiners. They outlined the current standard in intensity care units (ICUs) for assessing the level of necessary sedation in adults with the use of an ordinal scoring system such as the motor activity and assessment scale (MAAS) or the Richmond agitation-sedation scale (RASS). The RASS includes the assessment of the level of sedation and agitation on a scale of 0-6 with 0 being unresponsive and 6 being dangerously agitated. However, these assessments of agitation and sedation are subjective and therefore limited in accuracy and prone to error. As a result, computer vision techniques, although at this point speculative, have the potential to quantify agitation in sedated ICU patients. It may also be useful in paraplegics where whole body movements is not available for monitoring with computer vision techniques. However, head motion and facial grimacing provide an alternative for quantifying patient sedation and agitation. They also indicated that in future research they will investigate the use of digital imaging and video of a patient's entire body movement and facial expression for the assessment of agitation and sedation in the ICU. In summary, this expert control system can be utilized within a decision support to provide closed-loop control for ICU sedation and analgesia and critical-care monitoring for life saving interventions.

The study by Zamzmi et al (74) focused on an automatic pain recognizing algorithm for neonates. Other research has only focused on pain recognition in adults. For neonates, pain treatment and dosage of analgesics can have a huge impact on their development. Current manual scales of pain measurement in neonates look at aspects including face, legs, activity, crying and ability to console. These measurements are inconsistent due to heavy observer bias. The data used here contained video files of acute pain stimulus like heel lancing, immunization, etc., The novel algorithm used in this paper is a cascaded CNN architecture with 72,593 training parameters. This algorithm classifies neonatal facial images into Pain and No-Pain classes and has performed better than ResNet-50 and hand-crafted methods such as Local Binary Pattern with an accuracy of 91% and an AUC score of 0.93. As described above, in other reports and followed by this report by Zamzmi et al (74), facial expression technology seems to be most impressive in neonates and children. This is the area where rapid developments can be achieved with clinical applications.

Table 3 summarizes the literature of pain research with application of machine learning and/or artificial intelligence for neonatal pain.

Table 3. *Reports of pain research with application of machine learning and/or artificial intelligence for neonatal pain assessment.*

Author/Year	**Data Analysis And Machine Learning Methods**	**Conclusion/Comments**
Gholami et al (99), 2010 Relevance vector machine learning for neonate pain intensity assessment using digital imaging. IEEE Trans Biomed Eng 2010; 57:1457-1466.	Machine Learning Algorithm	A hypotheses of pain presence and pain intensity in 26 neonates was established through facial expressions that were plugged into a ML algorithm. A limitation of this study is that all patients were of Caucasian race, which limits diversity of results.
Zamzmi et al (74), 2019 Pain assessment from facial expression: Neonatal Convolutional Neural Network (N-CNN). 2019 International Joint Conference on Neural Networks (IJCNN) Budapest, Hungary, 2019, pp 1-7.	Deep Learning & Machine Learning Algorithms	This study looks at the facial expressions and body movements of 31 neonates before, during and several moments after treatment with a painful stimulus. The feasibility of automatic pain recognition in neonates was proven with high accuracy using ML and DL.

Pain Assessment in Special Settings

Multiple studies assessed pain in special settings, including emergency department, ICU, and dental pain.

Vu et al (73) in their study used ML and Deep Learning methods to identify if the patient is in the presence of pain on admission at the Emergency Department. The data used here is 2000 adult patient notes that were taken by the nurses. Word embedding methods were used to extract useful data from the notes. Algorithms used were SVM, Random Forests, CNN and Gated Recurrent Unit (GRU) and a rule-based model. CNN and GRUs were used as the patient feature identifiers which are then passed to a multi-layer perceptron with a SoftMax activation. The algorithms classified the data into pain and no-pain classes. GRU performed the best with an accuracy and F1-score of 91% and 90.96% respectively. Results show that deep learning methods performed better than ML methods, both of which performed better than the rule-based method (defined by a senior emergency department nurse). This type of research could be used to perform clinical audits on large datasets or as a potential real-time clinical tool.

Kobayashi et al (76) focused on monitoring patients continuously in ICU for danger to their survival rate due to pain left untreated. The three ML algorithms used are random forest, SVM and LR. This is a continuous way of evaluating pain in a patient objectively and semi-automatically. Vital signs are used for indications of pain. Objective pain assessment was performed using Critical-Care Pain Observation Tool (CPOT) and the data consisted of 117,190 CPOT assessments taken from 11,507 patients. Data preprocessing consisted of noise removal, slicing data into time series and normalization. The output labels were considered as negative for CPOT values from 0 to 2 and positive for scores 3 and above. Oversampling was performed to reduce class imbalances. Random Forests algorithm showed the highest accuracy among all the algorithms with an AUROC score of 0.853. This study showed promising results for the use of vital signs for pain tracking.

Hu et al (93) used neuroimaging techniques for objective assessment and localization of pain. Stimulus given to the 21 patients was clinical dental pain through 20 hypersensitive tooth simulations. DL models used were Neural Networks (3,5,6 and 7 layer), RNN (long short-term memory). Data used here is neuroimaging data called Functional Near-Infrared Spectroscopy (fNIRS). Pain was assessed by two methods: 1) Pain/No Pain classes 2) Left/right side pain and no pain classes. Their framework integrated optical neuroimaging, augmented reality and neural network-based DL methods. Neuroimaging data is transferred to an augmented reality device (Microsoft HoloLens) to plot cortical activity onto a 3D brain model to help the clinician. Best performing models were the 3-layer neural network for the 2-class classification with an accuracy of 80.37% and the 6-layer neural network for the 3-class classification with an accuracy of 74.23%.

Yang et al (95) identified pain intensities in patients suffering from sickle cell disease using physiological measures such as oxygen saturation, systolic blood pressure, diastolic blood pressure, pulse, respiratory rate, temperature, and self-reported pain score. Missing values were filled out using an imputation method called multiple imputation. The analysis was done in two methods: intra-individual level and inter-individual method. Dataset was created using 40 patients from Duke University Hospital. ML models used were: Multinomial logistic regression (MLR), K-NN, SVMs and random forests. It was found that pain scores are not linearly related to the six vital signs, for that reason, a linear model was not used for the pain prediction. Intra-individual analysis results show SVM has achieved the highest accuracy of the 4 algorithms. Its accuracy varied between 0.377 to 0.800. SVM also performed the best when comparing the weighted F1 scores at 0.529. Inter-individual results show MLR to be the best performing algorithm with accuracy at 0.429. Of the available features, SpO2, systolic blood pressure, pulse and temperature significantly affected the pain prediction. Other pain scales have also been tested with MLR showing the best accuracy. Of the available pain scales, there was good performance for the 4-point rating scale. Limitations of the study were sample size and potential confounders to changes in vital signs (dehydration and infection).

Yang et al (79) studied continuous pain assessment on Sickle Cell Disease patients using Wearable sensor data. Ten readings from the wearable device were recorded and important features were extracted and selected. The wearable device had sensors for heart rate monitor, GSR sensor, skin temperature sensor, 3 axis accelerometer and 3-axis gyroscope. Feature selection was done using the Embedded method after evaluating 2 other methods (filters and wrappers). Four embedded methods were used in the form of least absolute shrinkage and selection operator (LASSO) regression, Elastic Net, random forest and SVM. 10-fold validation was performed on the entire dataset. Data bootstrapping was performed to control and keep track of the stability of the results. There is a strong correlation between pain score and the data from the wearable device which shows prediction of pain scores with high precision. The best performing algorithms were the feature ensemble methods which improved the robustness as well as the stability of the features selected with the stacked (combined models) model having a Pearson correlation coefficient of 0.618 and root mean square error (RMSE) of 1.526.

In summary, there are only a few studies assessing ML in special settings, including emergency department, ICU, and dental pain. As of now, the evidence seems to be preliminary even though it is emerging. Consequently, future research will be of assistance in the application of ML in clinical settings.

Table 4 summarizes the literature of pain research with application of machine learning and/or artificial intelligence for pain assessment in special settings.

Automated Measurements

In the study by Parthipan et al (96) the authors focused on predicting the increase or decrease of post-operative pain at three time points. One of the most widely used prodrug opioids (Hydrocodone) (Prodrug opioid is a drug which requires metabolism and chemical modification to exert their pharmacological effect) is inhibited by the presence of commonly prescribed antidepressant selective serotonin reuptake inhibitors (SSRIs). AUC taken as the evaluation metric. Their NLP algorithm identified depression with an F1 score of 0.95

Table 4. *Reports of pain research with application of machine learning and/or artificial intelligence for pain assessment in special settings.*

Author/Year	Data Analysis And Machine Learning Methods	Conclusion/Comments
Vu et al (73), 2019 Identifying patients with pain in emergency departments using conventional machine learning and deep learning. Australasian Language Technology Association (ALTA) Sydney, Australia, 2019, pp 111-119.	Machine Learning Deep Learning Models	Deep Learning models and Machine Learning classified emergency room patients into categories of Pain or No Pain to expedite appropriate treatment.
Kobayashi et al (76), 2021 Semi-automated tracking of pain in critical care patients using artificial intelligence: A retrospective observational study. Sci Rep 2021; 11:5229	Artificial Intelligence	In 11,527 patients, three machine learning methods were used to detect pain through vital signs during Critical-Care Pain Observation Tool assessment. This study proved the usefulness of vital signs for active pain assessment.
Hu et al (93), 2019 Feasibility of a real-time clinical augmented reality and artificial intelligence framework for pain detection and localization from the brain. J Med Internet Res 2019; 21:e13594.	Artificial Intelligence	CLARAi, an augmented reality and artificial intelligence mobile neuroimaging framework, was used on 21 patients to detect pain and study localization in the brain during real-time dental pain stimulation.
Yang et al (95), 2018 Improving pain management in patients with sickle cell disease from physiological measures using machine learning techniques. Smart Health (Amst) 2018; 7-8:48-59.	Machine Learning	The pain scores of patients with sickle cell disease were predicted from physiological measures, excluding medical information using machine learning algorithms. A limitation in this study is the lack of demographic and personal information that was used to generate each prediction model.
Yang et al (79), 2019 Continuous Pain assessment using ensemble feature selection from wearable sensor data. Proceedings (IEEE Int Conf Bioinformatics Biomed) 2019; 2019:569-576.	Machine Learning?	A wrist worn device was used to collect physiological and body movement sensor data to provide continuous pain assessment for sickle cell disease management. Machine learning techniques were used to select for particular pain indicative features.

against their manually annotated data of 300 randomly sampled clinical notes. The algorithm results were as follows: Mean AUC was 0.87, 0.81 and 0.69 for the time periods of discharge, 3-week, and 8-week. ML model used here is Elastic Net regularized regression using a 10-fold cross validation with data from EHR with codes based on ICD-9-CM codes and vital signs of patients. Preoperative pain, surgery type and opioid tolerance were the features found to be the strongest predictors for post-operative pain control. The NLP algorithm was used to identify depression in unstructured patient notes. The feature vector used was of 65 features. Three separate models were built for the three separate time periods. The paper concludes that the drug recommendations based on the patient's other drug doses isn't considered and should be for an effective treatment and the patients with depression medication should be given direct-acting drugs instead of prodrug opioids.

The research by Rahman et al (80) focused on fluctuations or variations of pain scores with respect to time and predicting future pain volatility levels of patients using the 'Manage My Pain' app. Models used are LR with ridge estimators, LR with LASSO, Random Forests and SVM. Stratified 5-fold cross validation was performed. K-means clustering was used for the clustering of patients into low and high pain volatility classification. First objective is to classify patients between low and high volatility and the second objective is to predict the future pain volatility of the patient. Dataset is made of 782 users of the app with 329,070 pain records in the dataset. Classes were balanced according to distribution of feature groups. Subsampling procedure was used to solve the problem of unbalanced classes between low and high volatile pain groups. The threshold of 1.6 was established and values scoring higher than 1.6 were deemed as high pain volatile patients. Initially, a clustering algorithm was used to classify the patients into low and high pain volatility groups and then algorithms were used to predict the pain volatility of the patient at a 6-month time-period. Random Forest achieved the best prediction accuracy of approximately 70%.

Erdoğan and Oğul (78) studied a ML approach for the pain assessment through vital signs. In this paper the authors mention the process of creating a new pain database called Medical Information Mart in Intensive Care (MIMIC) where the patient data from wearable sensors is collected and annotated in an ICU. Pain assessment is done by well-trained nursing staff. The data is a time-series type with 2 classes: Pain or no Pain and recorded over 8 hours before pain onset. The time points are divided into 3-, 6- and 8-hour groups. Statistical time-domain features from time-series vital signs are extracted from the data. Median, variance, mean, RMS, mean absolute deviation and interquartile range algorithms used to test were Adaboost, Multilayer perceptron, LogitBoost and Random Forest. Of all the algorithms, Random Forest performs the best in 2 of the 3 cases (6 and 8 hours) with an accuracy of around 69% and AUROC of around 0.7. For 3 hours, LogitBoost performed better with an accuracy of 75.4% and AUROC of 0.656.

Lee et al (81) focused on identifying pain intensity estimation using a phone camera for post-surgical pain. Post-surgical pain management is critical for a successful outcome. The camera output is captured in 2D and 3D facial key points. Data was collected using post-surgical patients and their pain intensity is estimated. To capture the patient's face, the phone was mounted to a holder that extended from a neck pillow. Comparison was made with DeepFaceLIFT, which is a two-stage hierarchical learning algorithm where the first stage takes facial key points as input and gives frame wise VAS pain intensities. The second stage takes estimated VAS scores as input and calculates various metrics (min, max, median, etc.,) using the Gaussian Process Model with radial-based function (RBF) kernel-automatic relevance determination (ARD) kernel to estimate VAS rating for the entire sequence of the video. In this paper, the ML model used is multiple instance support vector machine which is one type of Multiple Instance Learning (MIL) approach. Methods using MIL approaches are more accurate in terms of mean absolute error compared to DeepFaceLIFT method. Data was sampled into 3 different approaches: Random Sampling (randomly select 'k' frames from the sequence), Uniform Sampling (equally spaced 'k' frames from the sequence) and cluster-based sampling (all set of frames are similar ex: pain frames only, neutral only). The method showed promising results for automatic pain recognition without input from the humans. The MIL algorithm performed better than DeepFaceLIFT with AUC score of 0.71 for 2D key points and 0.75 AUC score for 3D key points.

Atee et al (83) studied application for pain recognition for patients who cannot give self-reports. Extracorporeal pulse activation technology (EPAT) tool was used to address this problem. EPAT uses AI to detect micro-expressions from facial analysis and provides objective and reproducible pain presence reports. The performance was compared to the Abbey Pain Scale (APS).

EPAT is a hybrid scale which uses automated facial pain rating and a questionnaire-based approach for other domains such as Voice, Movement, Behavior, Activity and Body. The score range in EPAT is 0 to 42 with corresponding bands of 'no pain', 'mild', 'moderate' and 'severe' pain intensities. Pearson correlation was used to check the EPAT with respect to APS. Agreements between the measurements were assessed using Cohen's Kappa statistic. The Pearson's correlation coefficient between APS and EPAT is 0.882 at rest and 0.894 with movement. The weighted kappa score demonstrated there was moderate to good reliability between EPAT and APS. Internal consistency was excellent overall for EPAT versus APS. The study was conducted with minimal interruption and standard care is believed to elicit nociceptive pain and relates well with real-world context. Patients with various types of dementia and pain diagnosis were covered.

The study by Lopez-Martinez et al (84) studied pain detection using fNIRS. This method would be particularly useful for patients who are unconscious during tissue damage while undergoing surgery. Previous research showed good accuracy in pain detection, but the apparatus was difficult to use when the patient was in the supine position. To solve that issue, the new method uses prefrontal signals. The first objective is to check the usage of fNIRS for pain identification. Next, usage of Prefrontal signals for the classification is studied and lastly, personalizing the ML model allows the model to fit individuals with better performance. Forty-three healthy patients were included in the dataset and the patients undergo 3 states: First is a resting state where the patient is comfortably seated in a chair; second is non-painful brush stimuli and the third is 2 levels of electrical impulses (low and high electrical shock). Feature extraction was done using 'Discretized continuous wavelet transform'. ML methods used are LR (L1, L2), SVM (linear Kernel, RBF kernel) for single task classification and Hierarchical Bayesian logistic regression (HBLR) for individualized classification. Multi-task learning was the ML method used for binary classification. SVM with RBF kernel got the best accuracy with 69% for single task classification while HBLR gave the best accuracy with a score of 81%.

The study by Al-Qerem (85) used a generative adversarial network (GAN) model along with an SVM algorithm to classify pain in patients using biosensor data. GANs are used for data augmentation by generating artificial data by training on real data. Type of GAN used here is a Least Square GAN. Important features were extracted from the available data using the Boruta algorithm. The dataset used was BioVid, and consisted of 85 healthy patients subjected to 4 levels of heat stimulus. The dataset included electromyography (EMG), Skin Conductance Level (SCL) and electrocardiogram (ECG) signals adding up to around 160 features. For the pain intensity classification, the SVM model predicts classes for pain intensities from 0 to 4. The accuracy metrics are as follows: 1) Using real data, all level classification accuracy was 38.6% while individual class accuracies were: 74.5,78,82.9 and 86.8%. 2) Using selected features, all level classification accuracy was 38.6% while individual class accuracies were: 77.7,87.4,83.4 and 87.7%. 3) Using real and augmented data, all level classification accuracy was 82.8% while individual class accuracies were: 87.9,89.4,92 and 94.5%. 4) Using selected features for both real and augmented data, all level classification accuracy was 82.5% while individual class accuracies were: 89.4,89.2,91.4 and 94%. The results showed a huge boost in performance in pain intensity classification with the help of data augmentation using GAN models.

Kong et al (86) studied using ML methods to skin conductance data. Thermal grill was used for inducing various levels of heat and cold and a wrist worn EDA device with Bluetooth was used to transmit data to a smartphone. The research used a modified metric called modified time-varying index of sympathetic activity (MTVSymp) which is derived from time-varying index of sympathetic activity (TVSymp). The MTVSymp and TVSymp are calculated from the EDA signals and VAS was used. Real time feature used is a modified TVSymp. Methods applied on the data are: Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis, SVM with Linear Kernel, RBF kernel, Polynomial kernel with order 3, LR and K-NN (K=5). Leave-one-Subject-out cross validation was used to validate the models. Of all the methods, LR and LDA gave the best results for TVSymp data and P-SVM (SVM with Polynomial kernel with order 3 gave the best results for MTVSymp data. The accuracy scores were 90% for TVSymp and 90% for MTVSymp.

Tsai et al (88) utilized a Stacked Bottleneck LSTM model. The goal was to identify pain intensity from voice during the conversation of the patient and the medical professional during an emergency triage. The novel architecture used here is a deep bottleneck layer using LSTM which forms a low-dimensional representation of input features. Data used here consisted of audio-video recordings from TV talk shows and Triage database, physiological measurements, and other clinical outcome data. NRS pain score was collected at initial triage and follow-up period. The NRS scores have been converted into categories based on range parameters. LSTM model was used to overcome the vanishing gradient problem found in RNNs. Vocal data was extracted into two types of data: Prosodic and Spectral features. Promising results were achieved using the novel LSTM autoencoder architecture for classification of pain scores into respective pain classes. Accuracy of 72.3% was achieved for binary classification while 54.2% was achieved for 3-class classification.

Naeini et al (94) described a method of using IoT devices to monitor a patient's pain states and it is focusing on a long-term application for patient pain monitoring. BioVid Pain Database is used by the ML algorithm to learn and predict the Pain class. Here the classification is done into 5 classes i.e., 0 to 4 where 0 represents the 'No Pain class'. The use of wearable smart devices is used to monitor the patient's current state and a lot of emphasis is given to the computation power and the accuracy of the whole ecosystem. There are two aspects taken into consideration: 1) Video from the patient extracted into features using OpenFace toolbox 2) The bio sensor readings such as EMG, ECG, PPG and GSR. These data points are all fused together and passed onto a ML algorithm like SVM and Random Forest. The whole algorithm was run on Raspberry Pi 3 and Nvidia Jetson Boards. There was analysis performed on the performance of the algorithm on these low powered devices and the loss of accuracy was calculated. Data preprocessing operations were also performed on the sensor data to remove noise and outliers. The result proves that a long-term monitoring system is a feasible solution to the Pain Monitoring application using Edge IoT devices in a real time scenario Accuracy in Two-class testing varied between 53% and 79%.

Table 5 summarizes the literature of pain research with application of machine learning and/or artificial intelligence for automated measurements.

Spinal Diagnosis

Chronic spinal pain is a complex and multifactorial phenomenon. Consequently, the high prevalence of chronic spinal pain, the numerous modalities of treatments applied in management of the problem, and the growing social and economic costs continue to influence medical decision making. Despite its commonality, both in primary care and tertiary care, it is often difficult to reach a definitive diagnosis of the origin of spinal pain. Interventional techniques, one of the common modalities provided in managing spinal pain, are based on the philosophy of a neurophysiologic basis, in that when present, a structural origin of pain is important, with or without coexisting psychosocial abnormalities and comorbid conditions. A major source of exponential growth in treatment modalities is the inherent difficulty in obtaining an accurate diagnosis. In the search of a diagnosis, an accurate or incorrect diagnosis may lead not only to expensive diagnostic ventures, but to treatment failures resulting in wasted healthcare dollars and delivery and diversion of essential healthcare resources. Fundamental to proper treatment is an accurate diagnosis which is based on the reliability of the test used to make the diagnosis. There are no universally accepted gold standards for the diagnosis of spinal pain, regardless of the suspected course. Despite these issues, interventional techniques, along with multiple other modalities, have been increasing in utilization (2-8). Apart from epidural interventions and facet joint interventions, interventional pain management also includes spinal cord stimulation and a multitude of other minimally invasive surgical procedures.

Based on the developments and improvements reported thus far in the surgical literature, and a few reports in the pain management literature, it is important to incorporate ML and AI in managing complex issues of chronic spinal pain management.

Self-report pain ratings, along with functional status measurements, have been the gold standards in clinical assessment of spine. These parameters are highly variable, inherently subjective in nature, and significantly influenced by multidimensional factors. Consequently, research focused on the development of quantitative objective predictors, alongside self-report, aid in the diagnosis, estimate of prognosis, and prediction of treatment efficacy with increasing importance in managing chronic pain, specifically spinal pain (1,77,110). Consequently, multivariate ML techniques have used neuroimaging data to propose brain signature for evoked experimental pain (110). However, neuroimaging-based pain prediction continues to be in the discovery phase and evolving and is limited to discrimination of brain activity patterns contrasting noxious stimulus evoked painful versus nonpainful states in healthy, pain-free individuals and estimation of experimental pain ratings (77). Multiple studies have been published in this regard with modest accuracy. Some have combined various aspects and built

Table 5. *Reports of pain research with application of machine learning and/or artificial intelligence for automated measurements.*

Author/Year	**Data Analysis And Machine Learning Methods**	**Conclusion/Comments**
Parthipan et al (96), 2019 Predicting inadequate postoperative pain management in depressed patients: A machine learning approach. PLoS One 2019; 14:e0210575.	Natural Language Processing	In this report, the authors focused on predicting increase or decrease of postoperative pain at 3 time points, using hydrocodone and inhibition by the presence of commonly prescribed antidepressant selective serotonin reuptake inhibitors. The study was performed in 300 randomly sampled clinical notes. They identified preoperative pain, surgery type and opioid tolerance as the strongest predictors for postoperative pain. The study concluded that the drug recommendations based on the patient's other drug doses is not considered and should be for an effective treatment and the patients with depression medication should be given direct-acting drugs instead of prodrug opioids, such as hydromorphone, instead of hydrocodone. This finding has significant implications in managing acute pain.
Rahman et al (80), 2018 Defining and predicting pain volatility in users of the manage my pain app: Analysis using data mining and machine learning methods. J Med Internet Res 2018; 20:e12001	Machine Learning	A new pain volatility measure was used to distinguish high and low volatility and to predict volatility in users of the Manage My Pain mobile phone application 6 months after signup. Profile information and medical history was used to create unique predictive models for each individual.
Erdoğan and Oğul (78), 2020 Objective pain assessment using vital signs. Procedia Comp Sci 2020; 170:947-952.	Machine Learning	A machine learning based model used vital signs from patients in the ICU as inputs to provide an output of potential pain to provide a more objective means of classification.
Lee et al (81), 2020 Pain intensity estimation from mobile video using 2D and 3D facial keypoints. CoRR 2020; 2006:12246.	Statistical Analysis	This paper employs a new technology that capitalizes on smartphones to capture 2D and 3D facial cues through video to provide pain intensity estimations.
Atee et al (83), 2017 Pain assessment in dementia: Evaluation of a point-of-care technological solution. J Alzheimers Dis 2017; 60:137-150.	Statistical Analysis	The efficacy of the electronic pain detection tool (ePAT) was measured in 40 aged patients with dementia and history of pain-related issues. The ePAT proved to be a reliable method to detect pain in noncommunicative patients using automated facial expression assessment.
Lopez-Martinez et al (84), 2019 Pain detection with fNIRS-measured brain signals: A personalized machine learning approach using the wavelet transform and Bayesian hierarchical modeling with Dirichlet process priors. 2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), 2019.	Multi-task Machine Learning	Functional near-infrared spectroscopy was used to detect evoked electrical pain from the prefrontal cortex while using multi-task machine learning to personalize pain assessments in a sample of potentially high pain variability.
Al-Qerem (85), 2020 An efficient machine-learning model based on data augmentation for pain intensity recognition. Egypt Inform J 2020; 21:241-257.	Machine Learning	A machine learning algorithm was developed to identify pain intensity using feature selection whilst incorporating real and augmented data.

Table 5 (cont.). ***Reports of pain research with application of machine learning and/or artificial intelligence for automated measurements.***

Author/Year	**Data Analysis And Machine Learning Methods**	**Conclusion/Comments**
Kong et al (86), 2020 Pain detection using a smartphone in real time. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) Montreal, CAN 2020, pp 4526-4529..	Machine Learning	A wrist-worn electrodermal activity device was used to transmit signals to a smartphone for objective real-time pain detection in 10 individuals subjected to painful stimuli on the skin.
Tsai et al (88), 2017 Embedding stacked bottleneck vocal features in a LSTM architecture for automatic pain level classification during emergency triage. 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII) San Antonio, TX, 2017, pp 313-318.	Unsupervised Deep learning algorithm	Machine learning was used to detect the pain levels of patients in emergency triage through prosodic features during vocal speech. This was one of the first works where vocal cues determined pain in place of facial/behavior changes.
Naeini et al (94), 2019 An edge-assisted and smart system for real-time pain monitoring. 2019 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE) 2019, Arlington, VA, pp 47-52.	Deep learning and Machine learning algorithms	A self-aware system for real-time long term pain monitoring shows promise in wearable devices since it decreases energy consumption and holds high accuracy.

multivariate ML models that learn from central and autonomic features, and then classify clinical pain states and predict pain intensity. These approaches have been focused on chronic low back pain as the mainstay, followed by a multitude of other painful conditions.

Robinson et al (106) used commonly reported ML algorithms to measure differences between "objective" neuroimaging data and "subjective" self-report (i.e., mood and pain intensity) in their ability to differentiate between individuals with an without chronic pain. To achieve this, they used structure magnetic resonance imaging (MRI) data 26 individuals divided into 12 healthy controls and 14 patients with fibromyalgia and processed to derive volumes from 56 brain regions per individual. VAS ratings for pain intensity and mood (i.e., anger, anxiety, depression, frustration, fear) were self-reported. Then, separate models representing brain volumes, mood ratings, and pain intensity ratings were estimated across several ML algorithms. What the results showed was a classification accuracy of brain volumes ranging from 53-76%, while mood and pain intensity ratings ranged from 79-96% and 83-96% respectively. The authors concluded that models derived from self-report data outperformed neuroimaging models by an average of 22%. Additionally, they reported that while neuroimaging provides useful insight for the understanding of the neural mechanisms underlying pain processing, self-report is accurate, reliable and continues to be clinically vital. Overall, this study sheds significant light on the evaluation of patients using ML, supplementing with neuroimaging as required.

Mohan et al (107) combined resting state functional connectivity obtained from the source-localized electroencephalography of 311 tinnitus patients and 264 controls and a K-fold cross-validation ML algorithm to establish a predictive model that can verify the presence of behaviorally specific, spatiotemporally overlapping subnetworks in tinnitus. Even when compared to physiologically a similar disorder such as chronic pain, this complex reorganization was found to be exclusive to tinnitus even though each behavioral symptom has its own oscillating signature. What happens is that each frequency-specific transmission of information to be carried between 2 brain regions by way of the same anatomical connection. Not only does this provide additional understanding of the efficient mechanisms of compensation of the brain in the presence of multisystem disorders, but because of the exclusivity of the prediction model is presents the possibility for an objective neural marker for tinnitus. The authors have attempted to explain and to understand the mechanism of information transfer between different brain areas, one of the most interesting questions in neuroscience. By not only providing a possible explanation for the mechanism of information transfer, but also in the identification of different neuropathologies, network theory has gained traction and is at the forefront of this field. Not only that, but the perception of phantom ringing in the ear, called tinnitus, similar to other neuropathologies has been shown to be accompanied by aberrant functional connectivity between different brain areas. There have been independent studies showing the specific groups of areas encode individual symptoms of tinnitus, but there has not been any study to demonstrate that tinnitus is the unifocal percept of identifiable subnetworks encoding different behavioral aspects. Overall, in this study, the authors were able to explain efficient compensation mechanism of the brain in the presence of multisymptom disorders. The ML algorithm utilized in this model may be clinically applicable.

Sing et al (108), in a retrospective review, identified the top 100 spine research topics over a 37 year period of spine journal publications with the goal of accessing recent advances in machine learning (ML) which is computer learning without explicit instructions. They found broad technological advances and hypothesized that topic modeling algorithms can be applied to large volumes of text to discover quantifiable trends and themes. The significances of increasing ("hot") or decreasing ("cold") topic population over time was evaluated using simple linear regression. Spine related research articles from spine journals were extracted over a period from 1978 to 2015 yielding 25,805 articles which were subsequently classified into 100 topics. The top 2 most published topics were "clinical, surgeons, guidelines, care" which was found in 25% of 496 articles, and "pain, back, low, treatment, chronic." The top 2 hot trends were: 1) disc, cervical, replacement, level, arthroplasty (+0.05%/yr, $P < 0.001$) and 2) minimally, invasive approach, technique (+0.05%/yr, $P < 0.001$). The authors concluded that topics discovered through latent Dirichlet allocation modeling represent unbiased meaningful themes relevant to spine care. They further opined "that topic dynamics can provide historical context and direction for further research for aspiring investigators and trainees interested in spine careers." This study underscores the importance of AI and ML in spine care with numerous publications and replications of the data. Meaningful clinical applications are being developed in chronic pain management.

Lötsch et al (109) assessed quantitative sensory testing response patterns using the clinically established QST battery (German Research Network or Neuropathic Pain). They assessed well established experimental models of heat hyperalgesia of the skin which consisted of local ultraviolet-B (UV-B) irradiation or the application of capsaicin. They also included the application of cold and mechanical stimuli and measured their results in 82 healthy patients using the QST battery. They analyzed 10 QST parameters using machine learning techniques and found statistically significant effects in 9 of those parameters with regards to the effects of the hypersensitization treatments. Using supervised machine learning analysis which was implemented as random forests followed by ABC analysis they found heat pain thresholds to be the QST parameter most relevantly affected. Decision tree analysis, however, indicated that UV-b also modulated sensitivity to cold. Also, unsupervised machine-learning techniques, "implemented as emergent self-organizing maps, hinted at subgroups responding to topical application of capsaicin. The distinction among subgroups was based on sensitivity to pressure pain, which could be attributed to sex differences with women being more sensitive than men." The authors concluded that while UV-B capsaicin share a major component of heat pain sensitization, they differ in their effects on QST parameter patterns in healthy patients, suggesting a lack of redundancy between these models. Lötsch et al (109) attempted to combine multiple nociceptive measures with complex high dimensional data. Their conclusions appear to be acute with regard to a lack of redundancy between multiple models. Overall, this study included 82 healthy patients using a variety of noxious stimuli. This data, with replication in pain patients, may be applicable clinically with further assessment.

Abdullah et al (87) utilized physical spinal data of 310 patients which consisted of 12 feature measurements with two class labels. Unsupervised ML method, PCA is used for identifying features which affect the detection of spinal abnormalities. The goal of the model here is to identify whether the patient has spinal abnormalities or not (Normal/Abnormal). PCA was used to identify and prioritize the best features according to their significance. The *p*-score was used to identify significance by filtering out features that had a p-score greater than 0.05. Random forest was used to classify the patients into their respective classes based on their features.The K-NN algorithm was used to find the cluster to which the patient belongs. From the analysis, it was found that the feature 'degree spondylolisthesis' was the most significant feature affecting spinal abnormality. K-NN algorithm performed the best with an accuracy of 85.32%, followed by random forest with a score of 79.57%.

Abdollahi et al (92) utilized Inertial Measurement Unit (IMU) sensors attached to the patient's trunk while they performed trunk flexion and extension movements and a balance board was used to record center of pressure. Data used here is time scaled IMU signals of patient movements to understand kinematics. Models used are K-Means clustering, SVMs and multi-layer perceptron. The goal of the model is to classify non-specific low back pain patients into low, medium, and high risk. High risk patients undergo additional physiotherapy sessions which drastically differs from the other two groups, so identifying the right risk category is important for timely targeted intervention, thereby improving therapeutic outcomes. STarT Back Screening Tool (SBST) was used as the ground truth labels. Of all the models, SVM performed the best with an Accuracy score of 75.4, Sensitivity of 72.5 and Specificity of 78.2.

Botvinik-Nezer et al (97) examined variability in the analysis of a single neuroimaging dataset by many different terms. Because of the increasingly complex and flexible data analysis workflows, they wanted to examine the effect of this flexibility on the results of functional magnetic imaging. To accomplish this, they asked 70 independent teams to analyze the same dataset, testing, the same 9 ex ante hypothesis 1. The fact that no 2 teams chose identical workflows exemplified the flexibility of analytical approaches. There were sizeable differences in the results of hypothesis tests as a result of this flexibility even for teams "whose statistical maps were highly correlated at intermediate stages of the analysis pipeline." Several aspects of analysis methodology were responsible for this sizeable variation in reported results. Importantly, a meta-analytical approached "aggregated information across teams yielded significant consensus in activated regions. Furthermore, prediction markets of researchers in the field revealed an overestimation of the likelihood of significant findings, even by researchers with direct knowledge of the dataset 2, 3, 4, 5." They concluded that analytical flexibility can have a significant impact on scientific conclusions as well as identify factors that may be related to the sizeable variability in the analysis of fMRI. Furthermore, they stress the importance of validating and sharing and reporting multiple analyses of the same data. Additionally, they discussed potential approaches for mitigation of issues related to analytical variability.

As described above, based on the philosophy of multivariate ML model using multimodal neuroimaging and autonomic metrics, Lee et al (77) studied the prediction of clinical pain using fMRI and autonomic metrics along with ML methods. The lower and higher pain states in the patients were generated by performing certain physical exercises. This data was then used with multimodal ML approaches and applied to classify and predict clinical pain intensity. Patients reported pain intensities between 0-100 where 0 represented no pain and 100 is most pain imaginable. Classification of clinical pain states (lower and higher pain states) was done using SVM with a linear kernel. Prediction of pain intensities was done using support vector regression as pain intensity is a continuous variable. This paper used data fusion techniques to combine data from different sources such as functional connectivity of the primary somatosensory cortical representation of the back (S1CONN), regional cerebral brain flow and high frequency heart rate variability (HFHRV). It was concluded that head motion did not significantly affect the prediction accuracy of pain intensity ratings. Here the regional cerebral brain flow is obtained from arterial spin labeling fMRI-capturing slowly varying state changes in activity across the brain. S1CONN was obtained from blood oxygenation-level dependent (BOLD) fMRI which captures the temporal coherence of the S1 representation of the lower back. HFHRV captures the altered autonomic outflow associated with clinical pain perception. Here the HFHRV was captured by using pulse signals from the patient's finger instead of ECG to avoid noise from the MRI scanner. A limitation with the study is the small sample set in this experiment with only 53 patients. The ML model was able to use the fusion of multimodal brain and autonomic markers to classifying pain states and pain intensity prediction. Results for the SVM classification between high and low clinical pain states was Accuracy = 81.13% and the AUC = 0.90. For pain intensity ratings prediction, the results were: Training (Pearson's Correlation coefficient (r) =-.52, RMSE = 20.51) and for testing (r = 0.63, RMSE = 16.69). Overall, the results of the study from Lee et al indicate significant progress in optimism for the application of the multivariate model to assess chronic pain. However, as the authors have stated, this field is still nascent and the multivariate predictive models should not be used in lieu of subjective clinical pain ratings, but rather in conjunction with and supporting clinical pain ratings. Further, the authors have also noted multiple limitations, including a small number of patients with 53 patients. With replication of the study results and publications, future applications may extend these models to multiple sampling visits within a longitudinal trial framework. Overall, ML approaches with traditional assessments will amplify the accuracy of spinal disease states.

Table 6 summarizes the literature of pain research with application of machine learning and/or artificial intelligence for spinal diagnosis.

Treatment Algorithms

Andres et al (59) sought to assess predictive clinical decision system using machine learning and imaging biomarkers based on the philosophy that chronic pain is correlated with alteration in brain structure and function. The selection process for ideal candidate for spinal cord stimulation therapy is based on functional variables analysis and pain evaluation scores. They described multiple difficulties involved in the initial selection of patients and the predictive analysis of the trial phase and the large rate of expense as one of the most important concerns in the analysis of the suitability of implanted candidates. The objective was to investigate the usefulness of imaging biomarkers. Overall, this study includes only 24 patients; however, 7 were classified in the responders group. More importantly, by combining clinical variables and significant imaging biomarkers the prediction increased diagnostic accuracy in the responders group from 29% to 96% of long-term success, which is an admirable success rate. Overall, this study also shows the importance of clinical decision system using ML and imaging biomarkers in patients with neurostimulation therapy; however, this does not preclude clinical observations and a decision making process.

Soin et al (61) performed a pilot study implementing a ML algorithm to use AI to diagnose spinal conditions in chronic pain settings. Algorithmic approaches have been utilized in interventional pain management over the years (3,4), however, there are no publications related to AI and ML. Consequently, the authors evaluated whether it is possible to use AI via ML algorithms to analyze specific data points and to predict the most likely diagnosis related to spinal pain, in a prospective, observational pilot study. In this study, a total of 246 consecutive patients with spinal pain were enrolled. Patients were given an iPad to complete a Google form with 85 specific data points including demographic information, type of pain, pain score, pain location, pain duration, and functional status scores. The data were then input into a decision tree ML software program that attempted to learn which data points were most likely to correspond to the practitioner-assigned diagnosis. These outcomes were then compared with the practitioner assigned diagnosis in the chart. The results showed that the majority of patients had average pain history of 2 years with the majority being women. The most common practitioner assigned diagnosis included lumbar radiculopathy and lumbar facet joint disease. Comparison of the software predicted the diagnosis based on reported symptoms with the practitioner-assigned diagnosis revealed that the software was accurate approximately 72% of the time. They concluded that the software predicted diagnosis, based on the data from patients with spinal pain, had an accuracy rate of 72% suggesting promise for augmented decision-making using AI in this setting. Soin et al's (61) study shows the emerging interest in ML and AI. The 72% diagnostic accuracy is reasonably impressive. This is similar to the results shown in multiple reports utilizing an algorithmic approach with clinic assessment and diagnostic blocks (3,4,111-113). In these assessments, Manchikanti et al (111) were able to identify a pain generator in 81% of the population, whereas, other evaluations also ranged in similar proportions. The multiple variations in the diagnostic approaches with a paradigm shift from an acute to a chronic pain model also has shown significant changes with increasing ability to diagnose painful conditions (112,113). While additional studies are needed to build on the current approaches using ML, future directions could include investigating the ability to categorize patients with spinal pain into subgroups using a broad range of biopsychosocial factors, including incorporation of objective data from patient-owned devices. Overall, this pilot study provides positive input into ML in the diagnostic realm of spinal disorders.

d'Hollosy et al (91) utilized ML algorithms to determine if patient data could be used for decision making for the selection of treatment for low back pain patients. Dataset consisted of patient reported data from a spine center in the form of a biopsychosocial

Table 6. ***Reports of pain research with application of machine learning and/or artificial intelligence for spinal diagnosis.***

Author/Year	**Data Analysis And Machine Learning Methods**	**Conclusion/Comments**
Robinson et al (106), 2015 Comparison of machine classification algorithms for fibromyalgia: Neuroimages versus self-report. J Pain 2015; 16:472-477.	Machine Learning	This study compared the accuracies of objective neuroimaging data versus subjective self-reports in the classification of chronic pain to conclude that self-reports continue to be clinically reliable diagnostic tools.
Mohan et al (107), 2017 Evidence for behaviorally segregated, spatiotemporally overlapping subnetworks in phantom sound perception. Brain Connect 2017; 7:197-210.	Machine Learning	In a sample of 311 tinnitus patients, a machine learning algorithm developed an exclusive prediction model for the overlapping subnetworks producing this multisymptomatic neuropathology.
Lee et al (77), 2019 Machine learning-based prediction of clinical pain using multimodal neuroimaging and autonomic metrics. Pain 2019; 160:550-560.	Machine Learning	Multimodal brain and autonomic markers were used in a machine learning approach to classify pain thresholds in clinical exacerbation models in 53 individuals suffering from chronic low back pain.
Sing et al (108), 2017 Machine learning-based classification of 38 years of spine-related literature into 100 research topics. Spine (Phila Pa 1976) 2017; 42:863-870.	Machine Learning	Machine learning algorithms grouped 25,805 spine related research articles from 1978 to 2015 into 100 topics based on common themes.
Lötsch et al (109), 2018 Quantitative sensory testing response patterns to capsaicin- and ultraviolet-B-induced local skin hypersensitization in healthy patients: a machine-learned analysis. Pain 2018; 159:11-24.	Machine Learning	Biomedical data regarding ultraviolet-B light and capsaicin -induced hypersensitivity to local skin was clustered using machine learning to further the current understanding of nociception mechanisms.
Abdullah et al (87), 2018 Prediction of spinal abnormalities using machine learning techniques. 2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA) Kuching, MYS, 2018, pp 1-6).	Machine Learning	Degree spondylolisthesis was determined to be the greatest contributor to spinal abnormalities through supervised and unsupervised machine learning approaches.
Abdollahi et al (92), 2020 Using a motion sensor to categorize nonspecific low back pain patients: A machine learning approach. Sensors (Basel) 2020; 20:3600.	Machine Learning	In a sample of 94 patients, a wearable device attached to the trunk of each individual provided vital kinematic data to assess nonspecific low back pain.
Botvinik-Nezer et al (97), 2020 Variability in the analysis of a single neuroimaging dataset by many teams. Nature 2020; 582:84-88.	Statistical Analysis	A single neuroimaging dataset was uniquely analyzed by 70 independent teams to produce variable binary results. This study demonstrates how multi-verse analysis needs to converge on the same result in the future through more advanced tools.

questionnaire consisting of questions from 4 questionnaires. The dataset consisted of 287 input features. The algorithms are used to predict if the patient needed 'Pain Rehabilitation treatment' or 'Pain Surgery Treatment'. Only models with AUC scores of greater than 0.55 threshold were considered. Of the 25 ML models used, the best performing ones had an AUC score of 0.67 (BayesNet and Naive Bayes). The authors conclude that the AUC score needs to be closer to 0.72 to be able to use the algorithm in decision making clinical applications. Limitations of this study are the use of imbalanced data for the training process and the usage of cost sensitive learning to reduce the False-positive and False-negative errors/predictions. The results of this study were slightly inferior to the study by Soin et al (61). Further, they also confirmed that AUC score needs to be closer to 0.72 to be able to use the algorithm in decision making clinical applications. Soin et al (61) reported 72% accuracy. This study also adds to the growing literature which still is very small in managing chronic spinal pain patients using the ML algorithmic approach.

Segal et al (98) sought to develop a ML algorithm for the early detection of opioid use disorder. They gathered their data from a commercial claims database from January 1, 2006 to December 31, 2018, which included 10 million medical insurance claims from 55,000 patient records. The goal was to test the usefulness of ML in the creation of a prediction model and algorithm for early diagnosis of opioid use disorder. They put together 436 predictor candidates which they divided into 6 groups: demographics, chronic conditions, diagnosis and procedure features, medication features, medical costs, and episode counts. They used Word 2 Vec algorithm and the Gradient Boosting trees algorithm to analyze the data. They found that the c-statistic for the model was 0.959, with a sensitivity of 0.85 and a specificity of 0.882. The positive predictive value was 0.362 while the negative predictive value was 0.998. They found significant differences between positive opioid use disorder and negative opioid use disorder. The controls were in: the mean annual amount of opioid use days; the number of overlaps in opioid prescriptions per year; the mean annual benzodiazepine and muscle relaxant prescriptions. There were notable differences in: the count of intervertebral disc disorder-related complaints per year; post laminectomy syndrome diagnosed per year; and pain related disorder diagnosis per year. Additionally, there were significant differences in the episodes and costs categories. The authors concluded that this new algorithm offers a mean 14.4 months' reduction in the amount of time necessary to diagnose an opioid use disorder with potential savings in further morbidity, medical costs, addictions and mortality. As described by the authors, opioid use disorder effects 16 million people worldwide, but the diagnosis of opioid use disorder is commonly delayed or missed altogether. Thus, the algorithmic approach certainly has appeal and may be very useful. However, this has not been replicated and consequently it will be necessary to await replication and re-evaluation to determine its usefulness and appropriateness as an avenue for early diagnosis of opioid use disorder.

Table 7 summarizes the literature of pain research with application of machine learning and/or artificial intelligence for treatment algorithms.

Discussion

There has been a considerable increase in the number of papers published in the field of pain research using ML and deep learning methods. A search on Google Scholar using the keywords 'pain management,' AI has shown a constant growth of around 20% every year in the number of research publications. In this study, a wide range of algorithms and methods were reviewed in the field of pain research. The important findings and limitations present within the papers reviewed are described below.

Recent papers have seen an increase in using various data sources for pain research. Utilizing multiple modalities diminishes dependence on a single data source, providing a more comprehensive approach to understanding the patient's state (67,75,77,82).

The majority of papers reviewed focused on pain assessment and classification. A limited amount focused on pain management treatment because the accuracy of the models has not reached an acceptable level for real practice. The authors concluded that the AUC score is lower than it needs to be in order to use the algorithm in decision making clinical applications for low back pain (91). At the moment, AI can act as a support tool to assess problems in patients while the clinician holds the ropes for treatment recommendations.

Facial image analysis paves the way for an automated decision system for pain assessment (74,75,81-83,89,94). It provides an objective way for evaluating pain in patients through recognition of facial expressions. This type of analysis reduces observer bias from doctors accustomed to perceiving traditional displays of pain (74). Facial analysis has translated to point-of-

Table 7. *Reports of pain research with application of machine learning and/or artificial intelligence for treatment algorithms.*

Author/Year	Data Analysis And Machine Learning Methods	Conclusion/Comments
Andres et al (59), 2021 Predictive clinical decision system using machine learning and imaging biomarkers in patients with neurostimulation therapy: A pilot study. Pain Physician 2021; 24:E1279-E1290.	Supervised Machine Learning	Supervised machine learning was used to create a clinical decision support system that was capable of selecting patients that would benefit most from spinal cord stimulation to treat Failed back surgery syndrome.
Soin et al (61), 2021 A pilot study implementing a machine learning algorithm to use artificial intelligence to diagnose spinal conditions. Pain Physician 2021; in press.	Machine Learning	246 patients suffering from spinal pain entered medical and demographic information into a machine learning software with the capability of generating a diagnosis. The artificial diagnosis was then compared to the practitioner's diagnosis for a 72% rate of accuracy.
d'Hollosy et al (91), 2020 Applying machine learning on patient-reported data to model the selection of appropriate treatments for low back pain: A pilot study. In Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2020) 2020, pp 117-124.	Machine Learning	The study suggests that machine learning models on patient reported data can be used to assist physicians in the future with selection of treatment for patients with low back pain.
Segal et al (98), 2020 Development of a machine learning algorithm for early detection of opioid use disorder. Pharmacol Res Perspect 2020; 8:e00669.	Machine Learning	The development of a machine learning algorithm reduced clinical diagnostic time for opioid use disorder by an average of 14.4 months. The algorithm can expedite the onset of therapy while decreasing medical expenditures from system delays.

care applications such as EPAT that accurately "detect facial micro-expressions indicative of pain" (74). This is particularly useful for non-communicative patients such as those with dementia or Alzheimer's.

Physiological data has become another type of input for AI pain assessment. The physiological data plugged into algorithms can be collected through patient's vitals (68,69,78,79,85,86,94). Since vital signs can effectively be used in the evaluation of pain, the medical profession is on the cusp of redefining efficiency in terms of time and resources. At the same time, automated patient monitoring can be done using AI (76). Wearable devices have seen a spike in usage due to their cost effectiveness, ease of use, and constant real time data collection. These assets make the devices well suited for healthcare use to provide an objective monitoring of pain. Susam et al (70) demonstrated that pain identification could be accurately done using EDA data alone. EDA can be obtained through a single wearable sensor without the requirement of addition hardware. This simplicity translates to fast and efficient pain assessment. These devices are multifaceted; in the previous study signals in the epidermis layer were measured from the wrist, while Yang et al (79) measured physiological and body movement data from the same area. This paper proved that both of these factors were important in automatic pain estimation. Abdollahi et al (92) collected motion data through a wearable device located on the trunk of low back pain patients to categorize them into high, moderate, and low risk. The device prompted the division of patients into separate physiotherapy sessions based on their risk categories, which ultimately impacted recovery.

So far, all the studies discussed have involved AI with some form of direct patient contact, meaning that they have enhanced a portion of the pain management and evaluation process. The type of AI that we will be discussing now optimizes the analysis of clinical notes post checkup. NLP is capable of predicting inadequate postoperative pain management in depressed patients (96). By siphoning through the list of drugs prescribed to a patient, NLP can determine if incompatibilities exist between drug actions. Vu et al (73) addressed clinical notes in the emergency department to quickly prioritize patients in the most pain who required additional care. This diminishes waiting times for patients waiting for analgesia to be administered.

GANs are unsupervised learning methods where the models automatically detect patterns in the input data and generate new outputs based on the distributions of the input data. The usage of GANs for generation of synthetic data is very promising. The overall accuracy for pain classification increased from 38.6% to 82.8% when synthetic data was added using GANs, and the model training was performed (85). This is an important finding as pain datasets are usually limited and there exists huge variance in different patients. It is hard to acquire a lot of data per patient especially when the patient is being exposed to a painful stimulus. GANs can be used to produce data for datasets with an underrepresentation of a certain class, solving the class imbalance issue. GANs can also help in de-identification of patient records for the generation of medical data from training models. Also, medical imaging data is usually not available on a large scale due to the expensive process of medical annotation. GANs can help solve this problem with synthesis of more data.

Limitaions

With increasing interest in ML and AI in healthcare, certain limitations in use are granted. ML and AI both depend on a large availability of data to configure functions. This means that large pools of private healthcare information must be made available for these modalities to be accurate. The same data could yield different results based on the pipeline used. There is a need to validate the ML approach with different pipelines that can reaffirm the output of the model before deployment in clinical practice (97). Several studies are limited by their small sample sizes that skew the representation of a population. A smaller data set makes it much harder to isolate confounding variables; hence, future studies should make an effort to gather more data from a larger sample. However, a large pool of data can be even more useful when comprised of a diverse dataset (90). The second limitation stems from the fact that pain does not follow a universal standard; each person perceives the sensation in a unique way. A pain rating of 2 on the NRS for someone could be ranked as an 8 for another. A large variance in individual reactions to pain makes generalization by ML and AI more difficult (84) Future direction requires the development of individualized models to better tailor treatment for separate concerns of pain using methods like multi-task ML (84). The third limitation involves baseline classifications of pain. Studies at present tend to classify pain into two categories: no pain versus pain (67-74,76,78,80,86,93). This can only be utilized as a preliminary classification of pain for algorithms that have the capacity to use more advanced categories. It would be more beneficial for algorithms to use pain scales from 0-10 as input data because it would provide a more descriptive way of distinguishing pain on a spectrum versus a blanket yes or no. The presence of pain is known, but future research could benefit from heading towards a direction where pain is given a deductive value. The fourth limitation surrounds the collection of data in the digital world of healthcare. Working with EMR data is specific to distinct care locations and does not yield a comprehensive look at the patient data outside the care provider. In contrast, the insurance claims data shows a more detailed overall report of the patient's healthcare status and serves as a stronger foundation to generate more accurate models (98). However, the limitation of only using claims data is that it is restricted to billing elements in the patient's medical history and does not consider the clinical/medical context. As neither EMR nor claims data independently provides an all-inclusive look at a patient's medical status, fusion of both data types would reveal a finer understanding of the patient's medical history (98). The fifth limitation involves the lack of transparency of deep learning models. It is harder to understand how they reach conclusions because the inputs and features are not explicitly programmed by the human; deep learning models focus on self-chosen patterns in the data in order to identify tiers of significance. Without an online map outlining the processing steps of the algorithm, it can become difficult for healthcare providers to trust the 'Black Box' model since it harnesses the potential to generate inappropriate responses. This uncertainty can be avoided by utilizing more interpretable models (73) such as explainable AI, which humans can understand, manage, and trust.

The future direction of ML and AI application in healthcare requires updating algorithms to adapt to individual descriptions of pain based on medical history. Looking at verbal, social, and physical movements in videos would provide a temporal understanding of pain. The process would require you to process a sequence of images and then classify the subject's behavior into the corresponding pain bucket). Moving towards the analysis of sequential events houses greater potential than static images that could conceal pain indications.

Conclusion

Traditional pain assessment methods have a lot of limitations due to high variability in patient reported pain scores and perception of pain by different individuals. There is a need for generalized and automatic pain detection and recognition methods to objectively quantify pain. In this paper, state of the art ML and deep learning methods were analyzed in relation to pain management techniques. This paper provided the latest algorithms being used over the past two years. It provided the current state of the methods and sensor modalities while addressing shortcomings within the methods.

Author Contributions

The study was designed by JNN, AKV, and LM.

All authors contributed to preparation to the manuscript, reviewed, and approved the content with final version.

There was no external funding in preparation of this manuscript.

Acknowledgments

The authors wish to thank Bert Fellows, MA, Director Emeritus of Psychological Services at Pain Management Centers of America, for manuscript review, and Tonie M. Hatton and Diane E. Neihoff, transcriptionists, for their assistance in preparation of this manuscript. We would like to thank the editorial board of *Pain Physician* for review and criticism in improving the manuscript.

Author Affiliations

Mr. Nagireddi is a Machine Learning Engineer at Pain Management Centers of America, Evansville, IN.
njnischal@gmail.com.

Ms. Vyas is a Medical Research Associate at Pain Management Centers of America, Evansville, IN.
Amanyavyas@hotmail.com

Dr. Sanapati is Co-Founder and Director, Pain Management Centers of America, Evansville, IN, and Assistant Professor of Anesthesiology and Research, School of Medicine, LSU Health Sciences Center, New Orleans, LA, USA
msanapati@gmail.com

Dr. Soin is Medical Director, Ohio Pain Clinic, and Clinical Assistant Professor of Surgery at Wright State University, Dayton, OH, USA.
drsoin@gmail.com

Dr. Manchikanti is Co-Founder and Director, Pain Management Centers of America, Paducah, KY, Clinical Professor, Anesthesiology and Perioperative Medicine, University of Louisville, Louisville, KY, and Professor of Anesthesiology-Research, Department of Anesthesiology, School of Medicine, LSU Health Sciences Center, Shreveport, LA, USA.
drlm@thepainmd.com

References

1. Lötsch J, Ultsch A. Machine learning in pain research. *Pain* 2018; 159:623-630.
2. Manchikanti L, Centeno CJ, Atluri S, et al. Bone marrow concentrate (BMC) therapy in musculoskeletal disorders: Evidence-based policy position statement of American Society of Interventional Pain Physicians (ASIPP). *Pain Physician* 2020; 23:E85-E131.
3. Manchikanti L, Kaye AD, Soin A, et al. Comprehensive evidence-based guidelines for facet joint interventions in the management of chronic spinal pain: American Society of Interventional Pain Physicians (ASIPP) guidelines. *Pain Physician* 2020; 23:S1-S127.
4. Manchikanti L, Knezevic NN, Navani A, et al. Epidural interventions in the management of chronic spinal pain: American Society of Interventional Pain Physicians (ASIPP) comprehensive evidence-based guidelines. *Pain Physician* 2021; 24:S27-S208.
5. Manchikanti L, Pampati V, Soin A, et al. Trends of expenditures and utilization of facet joint interventions in fee-for-service (FFS) Medicare population from 2009-2018. *Pain Physician* 2020; 23:S129-S147.
6. Manchikanti L, Pampati V, Soin A, Sanapati MR, Kaye AD, Hirsch JA. Declining utilization and inflation-adjusted expenditures for epidural procedures in chronic spinal pain in the Medicare population. *Pain Physician* 2021; 24:1-15.
7. Manchikanti L, Pampati V, Vangala BP, et al. Spinal cord stimulation trends of utilization and expenditures in fee-for-service (FFS) Medicare population from 2009 to 2018. *Pain Physician* 2021; 24:293-308.
8. Manchikanti L, Sanapati J, Pampati V, Kaye AD, Hirsch JA. Utilization of vertebral augmentation procedures in the USA: A comparative analysis in Medicare fee-for-service population pre- and post-2009 trials. *Curr Pain Headache Rep* 2020; 24:22.
9. Institute of Medicine (IOM). *Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research.* The National Academies Press, Washington, DC, June 29, 2011.
10. U.S. Department of Health and Human Services. Pain Management Best Practices Inter-Agency Task Force.

Final Report on Pain Management Best Practices: Updates, Gaps, Inconsistencies, and Recommendations. May 9, 2019. Accessed 7/28/2021.www.hhs.gov/ash/advisory-committees/pain/reports/index.html

11. Manchikanti L, Singh V, Kaye AD, Hirsch JA. Lessons for better pain management in the future: Learning from the past. *Pain Ther* 2020; 9:373-391.
12. U.S. Burden of Disease Collaborators. The state of US health, 1990 – 2010: Burden of diseases, injuries, and risk factors. *JAMA* 2013; 310:591-608.
13. Dieleman JL, Baral R, Birger M, et al. US spending on personal health care and public health, 1996-2013. *JAMA* 2016; 316:2627-2646.
14. Dieleman JL, Cao J, Chapin A, et al. US health care spending by payer and health condition, 1996-2016. *JAMA* 2020; 323:863-884.
15. Keehan SP, Cuckler GA, Poisal JA, et al. National Health Expenditure Projections, 2019-28: Expected rebound in prices drives rising spending growth. *Health Aff (Millwood)* 2020; 39:704-714.
16. Manchikanti L, Vanaparthy R, Atluri S, Sachdeva H, Kaye AD, Hirsch JA. COVID-19 and the opioid epidemic: Two public health emergencies that intersect with chronic pain. *Pain Ther* 2021; 10:269-286.
17. Jha SS, Shah S, Calderon MD, Soin A, Manchikanti L. The effect of COVID-19 on interventional pain management practices: A physician burnout survey. *Pain Physician* 2020; 23:S271-S282.
18. Gharaei H, Diwan S. COVID-19 pandemic: Implications on interventional pain practice-a narrative review. *Pain Physician* 2020; 23:S311-S318.
19. Centers for Disease Control and Prevention. Increase in fatal drug overdoses across the United States driven by synthetic opioids before and during the COVID-19 pandemic. CDC Health Alert Network, December 17, 2020. Accessed 10/12/2021. https://emergency.cdc.gov/han/2020/han00438.asp?ACSTrackingID=USCDC_511-44961&ACSTrackingLabel=HAN%20438%20-%20General%20Public&deliveryName=USCDC_511-DM44961
20. Auyeung A, Wang H, Pirvulescu I, Knezevic NN. Impact of the COVID pandemic on chronic pain management. *Sib Med J* 2021; 2:197-212.
21. Merskey H, Bogduk N. Task Force on Taxonomy of the International Association for the Study of Pain. *Classification of Chronic Pain: Descriptions of Chronic Pain Syndromes and Definition of Pain Terms*. 2nd ed. IASP Press, Seattle, WA, 1994.
22. Manchikanti L, Falco FJE, Singh V, et al. An update of comprehensive evidence-based guidelines for interventional techniques of chronic spinal pain. Part I: Introduction and general considerations. *Pain Physician* 2013; 16:S1-S48.
23. Manchikanti L, Abdi S, Atluri S, et al. An update of comprehensive evidence-based guidelines for interventional techniques of chronic spinal pain: Part II: Guidance and recommendations. *Pain Physician* 2013; 16:S49-S283.
24. Fairbank JC, Pynsent PB. The Oswestry Disability Index. *Spine (Phila Pa 1976)* 2000; 25:2940-2952.
25. Mousavi SJ, Parnianpour M, Mehdian H, Montazeri A, Mobini B. The Oswestry Disability Index, the Roland-Morris Disability Questionnaire, and the Quebec Back Pain Disability Scale: Translation and validation studies of the Iranian versions. *Spine (Phila Pa 1976)* 2006; 31:E454-E459.
26. Cleland JA, Childs JD, Whitman JM. Psychometric properties of the Neck Disability Index and Numeric Pain Rating Scale in patients with mechanical neck pain. *Arch Phys Med Rehabil* 2008; 89:69-74.
27. EuroQol Research Foundation. EQ-5D-Y User Guide, Version 2.0, 2020. Accessed 10/6/2021. https://euroqol.org/publications/user-guides/
28. Pain Disability Index. Accessed 10/4/2021. www.nhms.org/Portals/96/Documents/Resources/Pain_Disability_Index.pdf
29. Sharma VK, Lepping P, Cummins AG, Copeland JR, Parhee R, Mottram P. The Global Mental Health Assessment Tool--Primary Care Version (GMHAT/PC). Development, reliability and validity. *World Psychiatry* 2004; 3:115-119.
30. Hays RD, Schalet BD, Spritzer KL, Cella D. Two-item PROMIS® global physical and mental health scales. *J Patient Rep Outcomes* 2017; 1:2.
31. Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. *Risk Manag Healthc Policy* 2011; 4:47-55.
32. Manchikanti L, Benyamin RM, Falco FJE, Hirsch JA. Metamorphosis of medicine in the United States: A carrot and stick policy of electronic medical records. *Pain Physician* 2014; 17:E671-E680.
33. Entzeridou E, Markopoulou E, Mollaki V. Public and physician's expectations and ethical concerns about electronic health record: Benefits outweigh risks except for information security. *Int J Med Inform* 2018; 110:98-107.
34. Tsou AY, Lehmann CU, Michel J, Solomon R, Possanza L, Gandhi T. Safe practices for copy and paste in the EHR. Systematic review, recommendations, and novel model for health IT collaboration. *Appl Clin Inform* 2017; 8:12-34.
35. Nicholas A. A better alternative to copy & paste. Note Swift Blog. Accessed 01/03/2022. https://noteswift.com/blog/better-alternative-copy-paste/
36. Alpert JS. The electronic medical record in 2016: Advantages and disadvantages. *Digit Med* 2016; 2:48-51.
37. Murphy K, Di Ruggiero E, Upshur R, et al. Artificial intelligence for good health: A scoping review of the ethics literature. *BMC Med Ethics* 2021; 22:14.
38. Schwab K. The Fourth Industrial Revolution: What it means and how to respond. *World Economic Forum*, January 14, 2016. Accessed 10/4/2021. www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/
39. AI in the UK: ready, willing and able? House of Lords, Select Committee on Artificial Intelligence, Report of Session 2017-19. Published by the Authority of the House of Lords. United Kingdom, 2018. Accessed 10/4/2021. https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf.
40. de Neufville R, Baum SD. Collective action on artificial intelligence: A primer and review. *Technol Soc* 2021; 66:101649.
41. Turing A. Computing machinery and intelligence. *Mind* 1950; LIX:433-460.
42. Samuel AL. Some studies in machine learning using the game of checkers. *IBM J Res Dev* 1959; 3:210-229
43. Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol Rev* 1958; 65:386-408
44. Ravi D, Wong C, Deligianni F, et al. Deep learning for health informatics. *IEEE J Biomed Health Inform* 2017; 21:4-21.
45. Using Apple Watch for Arrhythmia Detection. December 2020. Accessed 9/29/2021. www.apple.com/healthcare/docs/site/Apple_Watch_Arrhythmia_

Detection.pdf
46. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (WYSA) for digital mental well-being: Real-world data evaluation mixed-methods study. *JMIR Mhealth Uhealth* 2018; 6:e12106.
47. Hassan T, Seus D, Wollenberg J, et al. Automatic detection of pain from facial expressions: A survey. *IEEE Trans Pattern Anal Mach Intell* 2021; 43:1815-1831.
48. Wu M, Luo J. Wearable technology applications in healthcare: A literature review. *OJNI* 2019; 23(3). Accessed 9/29/2021. www.himss.org/resources/wearable-technology-applications-healthcare-literature-review
49. Lo WLA, Lei D, Li L, Huang DF, Tong KF. The perceived benefits of an artificial intelligence-embedded mobile app implementing evidence-based guidelines for the self-management of chronic neck and back pain: Observational study. *JMIR Mhealth Uhealth* 2018; 6:e198.
50. Lopez CD, Gazgalis A, Boddapati V, Shah RP, Cooper HJ, Geller JA. Artificial learning and machine learning decision guidance applications in total hip and knee arthroplasty: A systematic review. *Arthroplast Today* 2021; 11:103-112.
51. Wong D., Yip S. Machine learning classifies cancer. *Nature* 2018; 555:446.
52. Johnson KW, Torres Soto J, Glicksberg BS. Artificial intelligence in cardiology. *J Am Coll Cardiol* 2018; 71:2668.
53. Saber H, Somai M, Rajah GB, Scalzo F, Liebeskind DS. Predictive analytics and machine learning in stroke and neurovascular medicine. *Neurol Res* 2019; 41:681.
54. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP. Machine learning in cardiovascular medicine: Are we there yet? *Heart* 2018; 104:1156.
55. Cabitza F, Locoro A, Banfi G. Machine learning in orthopedics: A literature review. *Front Bioeng Biotechnol* 2018; 6:75.
56. Rashidi P, Edwards DA, Tighe PJ. Primer on machine learning: Utilization of large data set analyses to individualize pain management. *Curr Opin Anaesthesiol* 2019; 32:653-660.
57. Anan T, Kajiki S, Oka H, Fujii T, Kawamata K, Mori K, Matsudaira K. Effects of an artificial intelligence-assisted health program on workers with neck/shoulder pain/stiffness and low back pain: Randomized controlled trial. *JMIR Mhealth Uhealth* 2021; 9:e27535.
58. Amann J, Blasimme A, Vayena E, Frey D, Madai VI; Precise 4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak* 2020; 20:310.
59. Andres JE, Ten-Esteve A, Harutyunyan A, et al. Predictive clinical decision system using machine learning and imaging biomarkers in patients with neurostimulation therapy: A pilot study. *Pain Physician* 2021; 24:E1279-E1290.
60. Saheb T, Saheb T, Carpenter DO. Mapping research strands of ethics of artificial intelligence in healthcare: A bibliometric and content analysis. *Comput Biol Med* 2021; 135:104660.
61. Soin A, Manchikanti L. A pilot study implementing a machine learning algorithm to use artificial intelligence to diagnose spinal conditions. *Pain Physician* 2021; in press.
62. Alhaug OK, Dolatowski FC, Solberg TK, Lønne G. Criteria for failure and worsening after surgery for lumbar spinal stenosis: A prospective national spine registry observational study. *Spine J* 2021; 21:1489-1496.
63. Bernstein J. Letter to the Editor: Can machine learning algorithms predict which patients will achieve minimally clinically important differences from total joint arthroplasty? *Clin Orthop Relat Res* 2020; 478:1374-1375.
64. Fontana MA, Lyman S, Sarker GK, Padgett DE, MacLean CH. Can machine learning algorithms predict which patients will achieve minimally clinically important differences from total joint arthroplasty? *Clin Orthop Relat Res* 2019; 477:1267-1279.
65. The National Uniform Claims Committee. Specialty Designation for Interventional Pain Management- 09. Accessed 11/5/2021. www.cms.hhs.gov/transmittals/Downloads/r1779b3.pdf
66. Bartlett MS, Littlewort GC, Frank MG, Lee K. Automatic decoding of facial movements reveals deceptive pain expressions. *Curr Biol* 2014; 24:738-743.
67. Prkachin KM, Solomon PE. The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain. *Pain* 2008; 139:267-274.
68. Zhao YL, Yang HT, Hansma PK, Petzold L. How much does it hurt: A deep learning framework for chronic pain score assessment. 2020 International Conference on Data Mining Workshops (ICDMW). Sorrento, Italy, 2020, pp 651-660.
69. Walecki R, Rudovic O, Pantic M, Pavlovic V, Cohn JF. A framework for joint estimation and guided annotation of facial action unit intensity. 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Las Vegas, NV, 2016, pp 1460-1468.
70. Susam BT, Akcakaya M, Nezamfar H, et al. Automated pain assessment using electrodermal activity data and machine learning. *Annu Int Conf IEEE Eng Med Biol Soc* 2018; 2018:372-375.
71. Baltrusaitis T, Zadeh A, Lim YC, Morency LP. OpenFace 2.0: Facial behavior analysis toolkit. 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). 2018, pp 59-66.
72. Fodeh SJ, Finch D, Bouayad L, et al. Classifying clinical notes with pain assessment using machine learning. *Med Biol Eng Comput* 2018; 56:1285-1292.
73. Vu T, Nguyen A, Brown N, Hughes J. Identifying patients with pain in emergency departments using conventional machine learning and deep learning. Australasian Language Technology Association (ALTA). Sydney, Australia, 2019, pp 111-119.
74. Zamzmi G, Paul R, Goldgof D, Kasturi R, Sun Y. Pain assessment from facial expression: Neonatal Convolutional Neural Network (N-CNN). 2019 International Joint Conference on Neural Networks (IJCNN). Budapest, Hungary, 2019, pp 1-7.
75. Haque MA, Bautista RB, Noroozi F, et al. Deep multimodal pain recognition: A database and comparison of spatio-temporal visual modalities. 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). Xi'an, China, 2018, pp 250-257.
76. Kobayashi N, Shiga T, Ikumi S, Watanabe K, Murakami H, Yamauchi M. Semi-automated tracking of pain in critical care patients using artificial intelligence: A retrospective observational study. *Sci Rep* 2021; 11:5229.
77. Lee J, Mawla I, Kim J, et al. Machine learning-based prediction of clinical pain using multimodal neuroimaging and autonomic metrics. *Pain* 2019; 160:550-560.
78. Erdoğan B, Oğul H. Objective pain assessment using vital signs. *Procedia Comp Sci* 2020; 170:947-952.
79. Yang F, Banerjee T, Panaggio MJ,

Abrams DM, Shah NR. Continuous Pain assessment using ensemble feature selection from wearable sensor data. *Proceedings (IEEE Int Conf Bioinformatics Biomed)* 2019; 2019:569-576.

80. Rahman QA, Janmohamed T, Pirbaglou M, et al. Defining and predicting pain volatility in users of the manage my pain app: Analysis using data mining and machine learning methods. *J Med Internet Res* 2018; 20:e12001.
81. Lee M, Kennedy L, Girgensohn A, et al. Pain intensity estimation from mobile video using 2D and 3D facial keypoints. *CoRR* 2020; 2006:12246.
82. Bargshady G, Zhou X, Deo RC, Soar J, Whittaker F, Wang H. Ensemble neural network approach detecting pain intensity from facial expressions. *Artif Intell Med* 2020; 109:101954.
83. Atee M, Hoti K, Parsons R, Hughes JD. Pain assessment in dementia: Evaluation of a point-of-care technological solution. *J Alzheimers Dis* 2017; 60:137-150.
84. Lopez-Martinez D, Peng K, Lee A, Borsook D, Picard R. Pain detection with fNIRS-measured brain signals: A personalized machine learning approach using the wavelet transform and Bayesian hierarchical modeling with Dirichlet process priors. 2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), 2019.
85. Al-Qerem A. An efficient machine-learning model based on data augmentation for pain intensity recognition. *Egypt Inform J* 2020; 21:241-257.
86. Kong Y, Posada-Quintero HF, Chon KH. Pain detection using a smartphone in real time. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Montreal, CAN 2020, pp 4526-4529.
87. Abdullah AA, Yaakob A, Ibrahim Z. Prediction of spinal abnormalities using machine learning techniques. 2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA). Kuching, MYS, 2018, pp 1-6).
88. Tsai F, Weng Y, Ng C, Lee C. Embedding stacked bottleneck vocal features in a LSTM architecture for automatic pain level classification during emergency triage. 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII). San Antonio, TX, 2017, pp 313-318.
89. Bargshady G, Zhou X, Soar J, Deo RC, Whittaker F, Wang H. Enhanced deep learning algorithm development to detect pain intensity from facial expression images. *Expert Syst Appl* 2020; 149:113305.
90. Santana AN, de Santana CN, Montoya P. Chronic pain diagnosis using machine learning, questionnaires, and QST: A sensitivity experiment. *Diagnostics (Basel)* 2020; 10:958.
91. d'Hollosy W, van Velsen, L, Poel M, et al. Applying machine learning on patient-reported data to model the selection of appropriate treatments for low back pain: A pilot study. In Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2020), 2020, pp 117-124.
92. Abdollahi M, Ashouri S, Abedi M, et al. Using a motion sensor to categorize nonspecific low back pain patients: A machine learning approach. *Sensors (Basel)* 2020; 20:3600.
93. Hu XS, Nascimento TD, Bender MC, et al. Feasibility of a real-time clinical augmented reality and artificial intelligence framework for pain detection and localization from the brain. *J Med Internet Res* 2019; 21:e13594.
94. Naeini EK, Shahhosseini S, Subramanian A, Yin T, Rahmani AM, Dutt N. An edge-assisted and smart system for real-time pain monitoring. 2019 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE). 2019, Arlington, VA, pp 47-52.
95. Yang F, Banerjee T, Narine K, Shah N. Improving pain management in patients with sickle cell disease from physiological measures using machine learning techniques. *Smart Health (Amst)* 2018; 7-8:48-59.
96. Parthipan A, Banerjee I, Humphreys K, et al. Predicting inadequate postoperative pain management in depressed patients: A machine learning approach. *PLoS One* 2019; 14:e0210575.
97. Botvinik-Nezer R, Holzmeister F, Camerer CF, et al. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* 2020; 582:84-88.
98. Segal Z, Radinsky K, Elad G, et al. Development of a machine learning algorithm for early detection of opioid use disorder. *Pharmacol Res Perspect* 2020; 8:e00669.
99. Gholami B, Haddad WM, Tannenbaum AR. Relevance vector machine learning for neonate pain intensity assessment using digital imaging. *IEEE Trans Biomed Eng* 2010; 57:1457-1466.
100. Chesler EJ, Wilson SG, Lariviere WR, Rodriguez-Zas SL, Mogil JS. Identification and ranking of genetic and laboratory environment factors influencing a behavioral trait, thermal nociception, via computational analysis of a large data archive. *Neurosci Biobehav Rev* 2002; 26:907-923.
101. Huang Y, Zheng H, Nugent C, et al. Feature selection and classification in supporting report-based self-management for people with chronic pain. *IEEE Trans Inf Technol Biomed* 2011; 15:54-61.
102. Bui DD, Zeng-Treitler Q. Learning regular expressions for clinical text classification. *J Am Med Inform Assoc* 2014; 21:850-857.
103. Meng H, Bianchi-Berthouze N. Affective state level recognition in naturalistic facial and vocal expressions. *IEEE Trans Cybern* 2014; 44:315-328.
104. Lötsch J, Dimova V, Hermens H, et al. Pattern of neuropathic pain induced by topical capsaicin application in healthy subjects. *Pain* 2015; 156:405-414.
105. Dimova V, Oertel BG, Kabakci G, et al. A more pessimistic life orientation is associated with experimental inducibility of a neuropathy-like pain pattern in healthy individuals. *J Pain* 2015; 16:791-800.
106. Robinson ME, O'Shea AM, Craggs JG, Price DD, Letzen JE, Staud R. Comparison of machine classification algorithms for fibromyalgia: Neuroimages versus self-report. *J Pain* 2015; 16:472-477.
107. Mohan A, Moreno N, Song JJ, De Ridder D, Vanneste S. Evidence for behaviorally segregated, spatiotemporally overlapping subnetworks in phantom sound perception. *Brain Connect* 2017; 7:197-210.
108. Sing DC, Metz LN, Dudli S. Machine learning-based classification of 38 years of spine-related literature into 100 research topics. *Spine (Phila Pa 1976)* 2017; 42:863-870.
109. Lötsch J, Geisslinger G, Heinemann S, Lerch F, Oertel BG, Ultsch A. Quantitative sensory testing response patterns to capsaicin- and ultraviolet-B-induced local skin hypersensitization in healthy subjects: a machine-learned analysis. *Pain* 2018; 159:11-24.

110. Karhade AV, Schwab JH. Introduction to The Spine Journal special issue on artificial intelligence and machine learning. *Spine J* 2021; 21:1601-1603.
111. Manchikanti L, Singh V, Pampati V, et al. Evaluation of the relative contributions of various structures in chronic low back pain. *Pain Physician* 2001; 4:308-316.
112. Manchikanti L, Kosanovic R, Pampati V, et al. Low back pain and diagnostic lumbar facet joint nerve blocks: Assessment of prevalence, false-positive rates, and a philosophical paradigm shift from an acute to a chronic pain model. *Pain Physician* 2020; 23:519-530.
113. Manchikanti L, Kosanovic R, Cash KA, et al. Assessment of prevalence of cervical facet joint pain with diagnostic cervical medial branch blocks: Analysis based on chronic pain model. *Pain Physician* 2020; 23:531-540.