How is artificial intelligence helping the diagnosis of pain?: A Systematic Review

Jonathan Hsueh

Introduction

Complex decision-making shrouded in uncertainty is at the core of emergency medical treatment. Visits to the emergency department force doctors and nurses to identify patients with life-threatening conditions from the ones with more common benign diagnoses (1). Pain is a very important contributor to hospital admissions, especially emergencies (2). Artificial intelligence has the opportunity to streamline this part of the medical field. Doctors seek to accurately diagnose the pain lest run risks of worsening symptoms, more frequent hospital visits, and costlier treatments. Doctors currently use self-assessment models to diagnose pain (Visual Analog Scale VAS). However, many complex factors affect the reliability of such models, including consciousness and perception of pain (3). Artificial Intelligence has improved rapidly over the past couple of years, opening new opportunities in different fields. It has shown promise in streamlining the diagnosis of pain for medical purposes. Currently, there are live facial analyses, studies of clinical data, and x-ray imaging that have popped up. Can AI help with the diagnosis of pain? Especially during COVID, nurses and doctors have become overworked and in short supply. AI models may be a solution to diagnosing pain more efficiently and effectively.

Materials and Methods

We used Google Scholar, ScienceDirect, Springer, and Oxford Academic to perform a literature review. To holistically review topics related to Artificial Intelligence and pain diagnosis, we conducted a search with the terms “Artificial Intelligence,” “facial recognition,” and “pain detection.” Our search’s search query was “artificial intelligence in pain detection.” 80+ papers were identified as being useful to the question. Next, include and exclude factors were considered, such as having been published within the past 5 years, concerning artificial intelligence models directly diagnosing pain conditions, and open-source artificial intelligence models. Our final conclusion was ~30 papers.

Findings

Most research papers got their data and sources from volunteers and established medical databases. Some common databases were the UNBC-McMaster database, the MIntPAIN database, and the BioVid database. These databases collected medical images concerning pain and pain expression. The AI models used a variety of pain scales. The most used ones were the Numerical Rating Scale (NRS), the Prkachin and Solomon Pain Intensity (PSPI) scale, and the Visual Analog Scale (VAS) (4). There were many Machine Learning models and classifiers that researchers used. The basic models were Convolutional Neural Networks (CNN), Support Vector Machines (SVM), Double Machine Learning (DML), Gaussian Support Vector Machines (SVM), K-Nearest Neighbor, Temporal Convolutional Network (TCN),
Linear discriminant analysis, Logistic regression and linear regression models, and Neural transfer Convolutional Neural Networks (NT-CNN) (5), (6). The studies researched the model’s capabilities in pain detection and pain intensity estimation. The mean accuracy for the detection of pain among the papers was 85.05%. The mean accuracy for the current pain intensity was 73.90%.

The Automatic Coding of the Facial Action Coding System (FACS) is useful for diagnosing pain (7). Developing a pre-trained Machine Learning (ML) model is useful for diagnosing pain. The researchers created an ML model using AI/ML and trained it on frames and video sequences of patients expressing pain (3).

Conclusion

This review confirms that AI/ML technologies can be used to detect pain through facial expressions at a high potential. Artificial intelligence could be a helpful tool in providing objective accurate measurements of pain intensity. It would support doctors and clinicians to make more informed decisions during rush hour emergency moments.

An issue of training an Artificial Intelligence model is the need for large amounts of data. There is a threshold for the amount of data a model must be trained with to be relatively reliable, and the model can be unreliable if it lacks too much data. The issue is that live reliable data is hard to obtain. Hospitals are reluctant to give this data due to ethical issues surrounding consent to taking photos of patients. To address the limitations, it is important to encourage the sharing of more diverse and complex publicly available data with researchers to improve the AI models. This data has to be acquired with appropriate ethical considerations and proper permissions. High-quality, up-to-date data makes facial recognition models reliable and robust. More research is required to expand artificial intelligence’s capabilities and test its performance in different pain settings. It needs to train on and learn all kinds of pain conditions, and in great quantity, to assess its full potential as an aid in clinical practice. And, the patient’s ethical considerations around privacy and algorithm biases must be addressed.
Introduction

Pain is all-consuming: when you have it, not much else matters, and there is nothing you can do about it. Pain is an unpleasant subjective experience caused by actual or potential tissue damage associated with complex neurological and psychosocial components (8).

Pain is very complex. There are many different mechanisms of pain that all give the same signals of an unpleasant subjective experience. Pain can be an early-warning physiological protective system, essential to detect and minimize noxious stimuli. Pain also represents a heightened sensory sensitivity stimulating the repair of unavoidable tissue damage by discouraging physical contact in that area. Consequently, the pain felt can represent two completely different issues, a precursor to tissue harm or an alarm that actual tissue damage has occurred (9).

The crux of the issue is that distinguishing the correct diagnosis for pain is difficult and time-consuming. The most intuitive answer is self-reporting models, as they are highly individualized and dependent on the individual’s perception. Medical literature currently provides several useful pain scoring systems: the Visual Analogue Scale (VAS), the Verbal Rating Scale (VRS), and the Numerical Rating Scale (NRS) (10). These systems are reliable and appropriate, colloquially and professionally called the gold standard labeling (GSL) (7).

With all good things, there are significant shortcomings. Self-assessment models fall apart when individuals are either unconscious, cognitively impaired, or unable to articulate themselves verbally. This group accounts for a significant amount of emergency response patients. Our studies focus on cases where individuals are either unconscious, cognitively impaired, or unable to articulate themselves verbally (3). This follows a common trend of technological advancement that keeps the needs of those with disabilities in mind. For example, much advancement occurred in the architecture space after American Disabilities Act (ADA) guidelines were introduced in building construction. Technological advancement can follow this trend.

Diagnostic errors — that is, overlooking a disease, or diagnosing it erroneously or late — are a known occurrence in health care and can have tragic effects. Short-comings in pain diagnoses are a cause of misdiagnoses. Confirmed numbers of diagnostic errors are around 12 million, but they can top magnitudes of potentially 50 to 100 million cases. Misdiagnoses lead to an estimated 795,000 Americans becoming permanently disabled or passing away, according to a study in the British Medical Journal (11).

Costing thousands of dollars for every misdiagnosis, this can be devastating for patients with chronic pains or serious illnesses. Doctors have tools to address daily hospital challenges. Risk scores such as the Heart Score and early stress testing judge a patient’s risk of a disease based on inputted data. Traditionally, this was done by hand, but now artificial intelligence has been shown to streamline this process. By automating the processing of these algorithms, AI models can better utilize precious resources and save time with the power of linear regression and deep learning.
Artificial Intelligence can also transform the healthcare system by spearheading a revolutionary technology in healthcare: facial expression analysis. With AI advancing rapidly, there is evidence that more ambitious technology can be utilized and streamlined including facial recognition. Whereas 50 years ago AI was trapped in expensive corporate labs, now anyone can download an open-source AI model and make progress in the field. Programs such as Yolo, FaceReader, and other open-source facial recognition programs open the doors to new pain diagnosis tools (7). With the reach of the internet being nearly limitless, the advancement of facial expression models is only limited by imagination. This abundance of knowledge and easy access to quality resources is one of the greatest achievements of the digital age.

Our study aims to explore the current state of AI-assisted pain detection using both facial expressions and the automation of present risk scores. Our objectives are as follows.

1. Summarize the advancements of research in this field, including prominent methods used by the majority of researchers.
2. Identify and discuss potential implications and challenges of deploying the technology into the healthcare system.
3. Highlight research gaps and propose areas for future work.
Methods

In conducting our literature review, we systematically utilized major academic databases, including Google Scholar, ScienceDirect, Springer, and Oxford Academic.

We initiated the search with the terms “Artificial Intelligence,” “facial recognition,” and “pain detection.” Our search query was “artificial intelligence in pain detection,” identifying over 80 relevant papers. We applied an exclusion criterion based on publication date, eliminating papers published more than 5 years ago, ultimately reducing the selection to approximately 70 papers.

In the second phase, we shifted our focus to the specific topics of artificial intelligence and pain conditions. Papers were categorized into two groups, one addressing the advancement of facial recognition models and the other concerning clinical data analysis. Approximately 30 papers that did not align with our specific focus were excluded. Our final conclusion was ~40 papers.

In summary, our methodological approach to the literature review maintained a systematic and rigorous process, with careful paper selection in line with academic standards and research objectives.
Results

In total, we examined ~40 papers for the results and observed the attributes and characteristics of each study, to gain a better comprehensive understanding of the field.

Most research papers got their data and sources from volunteers and established medical databases. The majority of papers used medical databases dated from the 2000s to around 2018. Some common databases were the UNBC-McMaster database, the MIntPAIN database, and the BioVid database. These databases collected medical images concerning pain and pain expression. Live data was collected through volunteers and patients at hospitals.

Some of the data gathered was consensual, while others were more secretive.

*Table: Summary of studies assessing the use of AI to detect pain through facial expressions (12).*

The studies observed a wide variety of pain settings. These include postoperative pain, self-identified shoulder pain, cold pressor-induced pain, heat-induced pain, electrical-induced pain, (faked) states of acute illness, and chest pain. The majority of studies focused on self-identified shoulder pain and chest pain. Most of the data from databases were images of real pain expressions, while much of the data from volunteers and patients were faked illnesses and expressions.

The AI models used a variety of pain scales. The most used ones were the Numerical Rating Scale (NRS), the Prkachin and Solomon Pain Intensity (PSPI) scale, and the Visual Analog Scale (VAS) (4-27). Some researchers modified the NRS and categorized it into 4 pain intensities (no pain, low, medium, and severe) instead of the original version (15). Some researchers created their own pain stimuli-dependent assessments and stimuli-based pain levels (0-4) (25), and they tested their models based on those assessments. One study used lipopolysaccharide (LPS) which triggers an immune response resembling acute illness (27).

There were many Machine Learning models and classifiers that researchers used. The basic models were Convolutional Neural Networks (CNN), Support Vector Machines (SVM), Double Machine Learning (DML), Gaussian Support Vector Machines (SVM), K-Nearest Neighbor, Temporal Convolutional Network (TCN), Linear discriminant analysis, Logistic regression and linear regression models, and Neural transfer Convolutional Neural Networks (NT-CNN). Many studies had modified versions of the base models. These included
Convolutional Neural Network - Long short-term memory (CNN - LSTM), 2-CNNs with sample weighting (complementing each other), Convolutional Neural Network - Recurrent neural network (CNN - RNN), Double Machine Learning (DML) combined with Support Vector Machine (SVM), Hybrid Convolutional Neural Network (CNN) - bidirectional Long Short-term Memory (LSTM), and Neural transfer convolutional Neural Network (NT-CNN) with four Convolutional Neural Networks (CNN) (4-27). Most of these models are created by the paper’s researchers specifically for the study.

The studies researched the model’s capabilities in pain detection and pain intensity estimation. The mean accuracy for the detection of pain among the papers was 85.05%. The mean accuracy for the current pain intensity was 73.90% (12).

Automatic Coding of FACS

The Automatic Coding of the Facial Action Coding System (FACS) is useful for diagnosing pain (7). The first step of the two-step process is to identify the patient in Action Units (AU) which are parametrical indicators of facial muscular activity. The next step is for the model to automatically detect these AUs, analyze the data, and calculate a result.

Pre-trained ‘Lab-grown’ ML Model

Developing a pre-trained Machine Learning (ML) model is useful for diagnosing pain (3). The researchers created an ML model using AI/ML and trained it on frames and video sequences of patients expressing pain. These include video surveillance and pain-induced facial expressions. The model scores the pain based on the Prkachin and Solomon Pain Intensity (PSPI) scale (3). The model utilizes computer vision, linear regression, and deep learning. Common indicators of pain the model found included brow lowering, eye closure, orbit tightening, and levator muscle contraction. Verbal expressions of pain include vocalizing pain, and nonverbal expressions of pain include withdrawal behavior, lower body posture, and facial expressions.

This image depicts video surveillance being used to capture facial expressions associated with pain, which are then analyzed by a computer system using machine learning to provide an accurate output of pain detection or intensity estimation. Created with BioRender.com
Discussion

**Automatic Coding of FACS**

The Automatic Coding of the Facial Action Coding System (FACS) is useful for many reasons (7). The software FaceReader 9.1, developed by Vicarious Perception Technologies B.V. Amsterdam and distributed by the company Noldus (Noldus Information Technology bv, Wageningen, The Netherlands), was the first commercially available, advanced automatic algorithm that allowed detecting single AUs and was the global market leader among AU detection algorithms (AUDA) at the time of this study (7). FaceReader promises to enable the automatic detection of Action Units (AU) which are the parametrical indicators of facial muscular activity in the Facial Action Coding System (FACS). The FACS is a well-established and well-renowned tool to understand facial expressions and translate them into potential pain experiences and expressions. In the two-step approaches, FaceReader 9.1 specifies the AUs used in the intermediate step which provides a form of transparency. And, it compares very well to other automatic pain recognition systems. There is great comparability between manual FACS coding and FaceReader’s automatic AU coding. As a result, it is favored to be very promising for the future (7).

**Pre-trained ‘Lab-grown’ ML Model**

Developing a pre-trained Machine Learning (ML) model is very common in this field. The vast majority of the papers we studied included a ‘Lab-grown’ ML model (3). The background of creating these models was to develop a more “objective” way of assessing pain, and the present need for skilled clinicians to calculate Action Units (AU) on the Facial Action Coding System (FACS) provides a challenge. The AI/ML model technology can interpret facial expressions. Facial expressions are advantageous in AI/ML because they provide relevant data in each video frame and change over time (3). This data is the perfect kind to train computer vision systems on. They can perform this operation automatically by training a system to recognize the facial expressions connected to pain. The first step for automating the detection of pain is to develop a pre-trained ML system. Supervised ML models are trained with large datasets labeled with the correct output. Computer vision, algorithms, and mathematical models recognize patterns through linear regression, k-nearest neighbor, and deep learning. The next step is to train the models unsupervised by giving them new data to generate categorizations. The model’s computer vision extracts the facial features from live video data and identifies pain-related patterns. The computer subsequently provides an estimation of the subject’s pain experience.

**Different AI Models**

FACS is a comprehensive anatomically based system that objectively measures multiple facial expressions and movements. The algorithm could successfully determine acute illness in someone based on face and body cues (28). The FACS 2.0 can automate the analysis of facial
expressions and make it possible to diagnose patients even quicker than the original FACS. It can diagnose Acute Myocardial Infarction (AMI) and All-Cause Mortality (ACM). In general, these models train physicians and patients to read the results of the ML algorithm. They require lots of monitoring, oversight, and so-called algorithm stewardship frameworks to ensure safety and proper results. A big focus for this field is to provide for patients not only in rush hour situations but also those from demographics who cannot verbally communicate or give a mentally capable answer. These include those who are unconscious, unable to verbally articulate themselves, or those who cannot think rationally. This field is a reinforcement of a large trend in innovation that addresses the needs of the physically challenged.

The YOLO model is a one-stage detection model (29). The mean average precision is the current benchmark metric used by the computer vision research community to evaluate the robustness and accuracy of an object detection model. The advantage of using the YOLO-v5 model is the use of real-time detection systems (29). This program uses the VDO file and monitoring screen detection. It can be used to detect medical conditions during routine clinician work. It is highly accurate. Another method is to observe past clinical data like blood pressure and heart rate to make a decision (30). This is a great pairing to facial analysis, but most models don’t incorporate both facial analysis and clinical gestalt. This combination can be great at treating emergency conditions such as sepsis, which causes around 7.5 percent of hospital admissions (31). Sepsis is a serious condition in which the body responds improperly to an infection and uses the immune system’s weapons against the body. Early recognition of acute illness is critical for timely initiation of treatment, and the combination of facial analysis and clinical gestalt can efficiently diagnose this condition.

Generally, AI models can provide many benefits. It allows for a more effective triage of chest pain patients based on severity and condition. It can make for early discharges that prevent emergency department crowding. Physicians benefit from a model that can provide objective and accurate measurements of pain intensity, to help successfully diagnose patients (32). In situations where the patient is unable to give a rational judgment, AI models can aid in the detection of serious pain. Current developments are leading towards the development of a fully automated, rapid, standardized, and objective model (that is efficient, fast, requiring no humans, standard and not biased, and based on appearance over subjective input). As for useful AI models, deep learning tools significantly improved general accuracy. The random forest model was one of the more accurate models in each study (2).
Ethics and Limitations

Facial Expression Unreliability

Facial expressions are cooperative social signals to communicate one’s genuinely felt emotions, but sometimes they are used to mislead others. Deceptions are a part of everyday life, and artificial intelligence models cannot account for this issue \(^{(14)}\). They may be controlled to gain considerable adaptive advantages, including social acceptance, deliberately manipulating others, and suppressing and dissembling emotional expressions. In the studies, voluntary facial control makes it difficult for observers to discern honest signals from controlled or falsified ones \(^{(14)}\).

Subjects show varied facial responses to pain. Some subjects showed a lack of facial responses to pain: they may have low pain sensitivity resulting in a high tolerance \(^{(3)}\). Others have more facial responses and a higher pain sensitivity. The variation in the population creates many outliers and creates the need to figure out whether to remove or keep certain outliers.

Automatic pain detection is challenging because it is complex, subjective, and subject to a variety of factors, such as an individual’s personality, social context, and past experiences. Some of these mechanical limitations include the presence of head motion and rotation that can significantly reduce the accuracy of the AI model’s ability to detect Action Units (AU) \(^{(28)}\). Medical conditions such as Parkinson’s, stroke, facial injury, or deformity can affect facial shape and mobility.

Ethical Concerns

There are pretty important ethical concerns. There is a potential for errors and inaccuracies in pain detection models. Solely relying on inaccurate models could lead to dangerous or inappropriate decisions, such as misdiagnosis and inappropriate treatment \(^{(1)}\). Misdiagnosing certain severe conditions based on an inaccurate pain detection model may lead to low-quality care or prompt unnecessary surgery or medication \(^{(3)}\). These instances can erode trust between patients and doctors, and they can result in significant legal and financial implications.

There are concerns regarding patient privacy and autonomy. Patients must be provided informed consent before they can receive facial analysis, and they are allowed to refuse facial analysis. Up-to-date and diverse databases are scarce and outdated. This limits the development of reliable and widely generalizable AI models that detect pain \(^{(3)}\). There is a great lack of reliable real-world data to train AI models. As these programs require vast swathes of reliable images and videos, this limits the programs’ accuracy \(^{(3)}\). Additionally, algorithms might be trained for particular demographics, and they may be trained less for minority groups. Joy Buolamwini coined this as the “coded gaze” \(^{(33)}\). This could potentially further marginalize already vulnerable groups.
Conclusions

This review confirms that AI/ML technologies can be used to detect pain through facial expressions at a high potential. The rules also indicate that these models can accurately detect and quantify pain through facial expressions, and they often outperform human observers in pain assessment and speed. Therefore, artificial intelligence could be a helpful tool in providing objective accurate measurements of pain intensity. It would support doctors and clinicians to make more informed decisions during rush hour emergency moments.

To address the limitations, it is important to encourage the sharing of more diverse and complex publicly available data with researchers to improve the AI models. This data has to be acquired with appropriate ethical considerations and proper permissions. High-quality, up-to-date data makes facial recognition models reliable and robust. Additionally, well-designed randomized control trials are necessary to determine the reliability and generalizability of automated pain detection in real-life clinical scenarios. These trials can determine if AI/ML models are appropriate in real-life settings with variability and diverse patient medical conditions.

More research is required to expand artificial intelligence’s capabilities and test its performance in different pain settings. It needs to train on and learn all kinds of pain conditions, and in great quantity, to assess its full potential as an aid in clinical practice. Furthermore, patients’ preferences and satisfaction regarding the usage of AI in medical settings should be explored. At the end of the day, the AI models serve the patients, so patient’s ethical considerations around privacy and algorithm biases must be addressed.
References


www.sciencedirect.com/science/article/abs/pii/S0169260718316936?casa_token=SgZQCscCHCgAAAAA%3Ag0lw9tO8DqW_6HXwnUA3KWy8cpqtGdgpnsb6UWA0WeFSUMRQ9Jv5wl5HT4a2xamtDfXaqULIQ.


www.painphysicianjournal.com/current/pdf?article=NzQyOQ%3D%3D&journal=142.


