Computer Methods and Programs in Biomedicine 231 (2023) 107365

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

# Automatic assessment of pain based on deep learning methods: A systematic review

Stefanos Gkikas[a,b,*], Manolis Tsiknakis[a,b]

[a] Department of Electrical and Computer Engineering, Hellenic Mediterranean University, Estavromenos, Heraklion, 71410, Greece
[b] Computational BioMedicine Laboratory, Institute of Computer Science, Foundation for Research & Technology-Hellas, Vassilika Vouton, Heraklion, 70013, Greece

ARTICLE INFO

*Article history:*
Received 25 May 2022
Revised 6 January 2023
Accepted 21 January 2023

*Keywords:*
Pain recognition
Affective computing
Machine learning
Facial expression
Biosignals

ABSTRACT

*Background and Objective:* The automatic assessment of pain is vital in designing optimal pain management interventions focused on reducing suffering and preventing the functional decline of patients. In recent years, there has been a surge in the adoption of deep learning algorithms by researchers attempting to encode the multidimensional nature of pain into meaningful features. This systematic review aims to discuss the models, the methods, and the types of data employed in establishing the foundation of a deep learning-based automatic pain assessment system.

*Methods:* The systematic review was conducted by identifying original studies searching digital libraries, namely Scopus, IEEE Xplore, and ACM Digital Library. Inclusion and exclusion criteria were applied to retrieve and select those of interest, published until December 2021.

*Results:* A total of one hundred and ten publications were identified and categorized by the number of information channels used (unimodal versus multimodal approaches) and whether the temporal dimension was also used.

*Conclusions:* This review demonstrates the importance of multimodal approaches for automatic pain estimation, especially in clinical settings, and also reveals that significant improvements are observed when the temporal exploitation of modalities is included. It provides suggestions regarding better-performing deep architectures and learning methods. Also, it provides suggestions for adopting robust evaluation protocols and interpretation methods to provide objective and comprehensible results. Furthermore, the review presents the limitations of the available pain databases for optimally supporting deep learning model development, validation, and application as decision-support tools in real-life scenarios.

## 1. Introduction

Pain, according to the International Association for the Study of Pain (IASP), is described as *"an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such tissue damage"* [1, p. 250]. Pain is a highly prevalent and manifold condition [2]. According to the Global Burden of Disease (GBD) study, pain is the number one cause of years lived with disability (YLD) [3]. The main types of pain are acute and chronic. Their primary difference has to do with the duration of the sensation of pain; pain is considered acute when it is present for less than three months and is probably accompanied by apparent physiological damage, while it is considered chronic when it progresses from an acute to a chronic state and persists beyond the healing process [4]. Chronic pain has several variations when the temporal dimension is taken into consideration, such as chronic-recurrent (*e.g.*, migraine headache) or chronic-continuous (*e.g.*, low back pain) [5]. Pain is a serious issue that concerns not only individuals but also society as a whole. Every day, people of all ages experience pain, either due to an accident, an illness, or even during treatment, and it is the most frequent reason for a physician visit. Both acute and chronic pain constitute clinical, economic, and social constraints [6]. The Institute of Medicine states that more than 100 million Americans suffer from chronic pain [7], and the lost productivity is estimated at 61.2 billion dollars per year [8]. The National Institute of Health (NIH) disclosed that

[*] Corresponding author.
*E-mail addresses:* gkikas@ics.forth.gr (S. Gkikas), tsiknaki@ics.forth.gr (M. Tsiknakis).

https://doi.org/10.1016/j.cmpb.2023.107365

the total cost of persistent pain ($560-$635 billion) significantly exceeds the cost of other major diseases, including cardiovascular ($309 billion) and neoplasms ($243 billion) [9]. Besides the direct consequences of pain in a patient's life, there are various collateral negative impacts related to opioids, drug overuse, addiction, poor social relationships, and psychological diseases [6]. Accurate pain measurement facilitates early diagnosis, monitoring of disease progression, and evaluation of therapeutic efficacy; thus is critical for the management of chronic pain.

For all the aforementioned reasons, the objective assessment of pain is required to deliver appropriate care to people suffering. A serious case concerning the healthcare community is vulnerable groups of people who cannot communicate and express their pain directly. Such groups of people are infants, young children, people with mental illness, and the elderly. Usually, in most cases involving people who cannot express their pain, caregivers or family members observe the behavioral or physiological responses through which they infer the presence or absence of pain. The problems with such an approach are twofold [10]. The first relates to the fact that continuous observation of the patient throughout the day is only possible by employing technology-based solutions. The second relates to accuracy, *i.e.,* whether the observation and drawing of conclusions are objective and correct. There are situations in which the observer, either due to inadequate training or personal biases, is not able to assess appropriately and sufficiently the pain event that the patient is experiencing [11]. In addition, social and interpersonal relationships influence the judgment and the expression of the person, those who evaluate, and those who perceive pain [12]. For the above reasons, significant research is focused on the development of automatic pain recognition systems that can recognize the existence of pain and its intensity, analyzing physiological and behavioral responses. In the last decade, artificial intelligence (AI) researchers have focused on creating models and algorithms capable of endowing machines with cognitive capabilities to recognize emotions and sentiments, such as pain. Especially in recent years, with the development of deep learning methods, many researchers are using such approaches for automated pain assessment, as the results are often superior to those of classical machine learning techniques.

Consequently, evaluating the current primary research studies adopting deep learning methods is necessary. According to [13], the aggregation of empirical results can be performed with systematic literature reviews (SLRs) addressing specific research questions or with systematic mapping reviews identifying all the relevant research articles on a specific topic. The present work is an SLR to report on the achievements of existing approaches in automatic pain assessment based on deep learning methods. We will fill the gap in the existing comprehensive reviews and provide insights about the techniques and strategies for future automatic pain recognition systems to practitioners and researchers in this scientific field.

### 1.1. Related work

Prior to making the final decision to conduct an SLR, the literature was reviewed, and existing SLRs on pain assessment were identified and assessed. The following were observed. In 2009, the first review [14] on automatic pain assessment was published, which does not include papers based on deep learning, as the actual implementations of deep architectures started in 2012. Zamzi et al. [15] conducted a review of automatic pain assessment, which focused exclusively on infants and did not reference deep learning methods. In 2018, Chen et al. [16] presented a review focusing on automated pain detection approaches using the Facial Action Coding System (FACS). The authors of this review also report only a limited number (*i.e.,* three) of publications using deep learning methods. A year later, *i.e.,* in 2019, Hassan et al. [17] presented a similar review in which only seven papers reporting the use of deep learning methods were included. The same year, Werner et al. [18] presented their results on pain assessment without any constraints on the modalities used or the age of subjects. The papers reporting deep learning methods are again less than ten. In 2020 Al-Eidan et al. [19] published the first SLR on pain assessment and deep learning approaches entitled "Deep-Learning-Based Models for Pain Recognition: A Systematic Review", which includes fifteen papers. This review has, in our view, significant limitations and incorrect information. We believe that a number of papers analyzed are not relevant, and there is, in our view, confusion between the terms "neural networks" and "deep learning" since the presence of the first in a study does not necessarily imply the existence of the second. Specifically, although the referenced studies [3,17] report the use of neural network approaches, they do not present any evidence regarding the adoption of deep learning methods. At the same time, in [16], the authors explicitly report the development of a neural network with two layers combined with handcrafted features, which is certainly not a deep learning method. Furthermore, the studies [15,19] focus on detecting protective movement behavior in patients suffering from chronic pain, which is a different research topic.

As a result, although several reviews and SLRs have been published on automatic pain assessment, they do not focus exclusively or properly on deep learning methods. The present SLR attempts to fill this gap in the literature by providing a comprehensive systematic review of deep learning methods employed for automatic pain assessment.

### 1.2. Pain

Pain is explained as an unpleasant noxious stimulus originating from the peripheral nervous system and transferred through the spinal cord, followed by the physiological sensation of it. It is a complex biopsychosocial sensation that emerges from the synergy of neuroanatomic and neurochemical systems combined with cognitive and affective processes [20]. For the above reason, Williams and Craig [21] proposed an updated definition of pain: *"Pain is a distressing experience associated with actual or potential tissue damage with sensory, emotional, cognitive and social components".*

Three main variables characterize pain; severity, duration, and distribution [5]. Pain severity is the most visible element, with low, moderate, or high intensity. The second variable is the duration of pain; as previously mentioned, the two primary types of pain are acute and chronic. The last variable is the distribution of pain. Pain distribution is one of the typical factors used for the clinical assessment of patients with chronic pain. In order to effectively manage the pain situation of an individual, it is necessary to assess its presence and intensity. In clinical settings, the gold standard is the self-report, where the person describes the intensity or/and the region where the pain occurs. There are numerous self-report scales related to adults, children, and elders, such as the visual analogue scale (VAS) [22] and the verbal rating scale (VRS) [23]. Additionally, there are observational-based scales where a third person evaluates the severity of the pain, *e.g.,* the Prkachin and Solomon pain intensity scale (PSPI) [24] and the neonatal/infant pain scale (NIPS) [25]. However, there is reported evidence that patients report high pain severity in order to provoke more aggressive treatment [26]. These types of incidents create uncertainties about the validity of the reported symptoms. Hence the objective measurement of pain intensity is clinically fundamental.

### 1.3. Modalities & hardware for automatic pain assessment

The development of an automatic pain assessment system requires the recording of the necessary input information channels.

**Table 1**
Pain Databases utilized in most studies.

| Database | Modality | Population | Annotation Granularity | Annotation Labels |
|---|---|---|---|---|
| **UNBC-McMaster** Shoulder Pain$^A$ [27] | RGB video of face | 25 adults with shoulder pain | Frame level<br>Sequence level | FACS<br>VAS, OPI |
| **BioVid**$^A$ [28] | RGB video of face, EDA, ECG, EMG | 87 healthy adults | Sequence level | stimulus (calibrated per person) |
| **MIntPAIN**$^A$ [29] | RGB-Depth-Thermal video of face | 20 healthy adults | Sequence level | stimulus (calibrated per person), VAS |
| **iCOPE**$^A$ [30] | RGB photographs of face | 26 healthy neonates | Frame level | pain, cry, rest, air puff, friction |
| **iCOPEvid**$^A$ [31] | Grayscale video of face | 49 neonates | Sequence level | pain, no pain |
| **NPAD-I**$^A$ [32] | RGB video of face & body, HR, SpO2, BP, NIRS | 36 healthy neonates & 9 neonates with tissue injured by surgery | Sequence level | NIPS, N-PASS |
| **APN-db**$^A$ [33] | RGB video of face | 112 healthy neonates | Sequence level | NFLAPS, NIPS, NFCS |
| **EmoPain**$^N$ [34] | video, audio, EMG, MoCap | 22 adults with chronic pack pain & 28 healthy adults | Sequence level | self-report, naive OPI |
| **SenseEmotion**$^N$ [35] | video of face, audio, EDA, ECG, EMG, RSP | 45 healthy adults | Sequence level | stimulus (calibrated per person) |
| **X-ITE**$^N$ [36] | RGB-Thermal video of face, RGB-Depth video of body, audio, EDA, ECG, EMG | 134 healthy adults | Sequence level | stimulus (calibrated per person) |

$^A$: Publicly available by request, complete or part of the dataset $^N$: Not yet available **Modality:** HR: heart rate SpO2: oxygen saturation rate BP: blood pressure NIRS: near-infrared spectroscopy MoCap: motion capture RSP: respiration rate EDA: electrodermal activity ECG: electrocardiogram EMG: electromyogram **Annotation Labels:** FACS: Facial Action Coding System VAS: visual analogue scale OPI: observer pain intensity NIPS: neonatal infant scale N-PASS: neonatal pain, agitation and sedation scale NFLAPS: neonatal face and limb acute pain scale NFCS: neonatal facial coding system

The information channels known as modalities are characterized as behavioral or physiological. A system is called unimodal if it consists of one modality or multimodal if multiple modalities are used.

The primary behavioral modalities include facial expressions, body movements, gestures, and audio. Several kinds of optical and light sensors can be exploited to record images or sequences of images (*i.e.,* videos) regarding facial and body expressions. Most researchers utilize color RGB cameras, while in other cases, depth and thermal camera sensors are employed to provide additional visual channels. In addition, motion capture sensors have been used for movement measurement, while microphones are the typical hardware choice for audio recording. Physiological modalities usually include biosignals capturing the electrical activity arising from tissues and organs. Numerous biosignal measurement methods have been used in order to assess pain; electrocardiography (ECG), electromyography (EMG), electrodermal activity (EDA), photoplethysmography (PPG), blood oxygen saturation (SpO2), near-infrared spectroscopy (NIRS), respiration rate and skin temperature. It is worth mentioning that multiple modalities can be measured by several sensors, *e.g.,* respiration rate can also be measured with strain sensors and cameras.

Beyond the sensors capturing the necessary input information, the hardware for the computation process is essential. Deep learning-based systems works in two phases; training and inference. The training phase is the most computationally intensive, and a graphics processing unit (GPU), even for efficient approaches, is necessary. For inference, the trained model is deployed, making predictions on novel data. Usually, this process takes place in a central processing unit (CPU), but the specific choice of hardware is related to several factors. For instance, in real-time scenarios where the latency is critical, the requirements are higher than in an offline approach, where estimations happen in a subsequent stage. Furthermore, the characteristics of the trained model, *i.e.,* floating point per second (FLOPS), and the number of operations must be considered too.

### 1.4. Pain databases

Data availability is essential to evaluate different methods and algorithms for automatic pain assessment. Table 1 summarizes the main databases used in the studies reviewed. In Fig. 1, we present the number of studies that utilized each database. Most studies employed one of the publicly available datasets, and several studies experimented with more than one dataset. A limited number of studies employed a private dataset, especially those focusing on the pain detection of neonates. The most used dataset is UNBC-McMaster Shoulder Pain Archive Database [27], followed by The BioVid Heat Pain Database [28]. The first consists of 200 facial videos of 25 participants suffering from shoulder pain, while the second combines facial videos and biopotentials of 90 healthy participants subjected to experimentally induced heat pain at four different intensity levels.

### 1.5. Structure of the paper

The review is organized as follows: in Section 1, we provide a brief analysis and assessment of existing SLRs and briefly describe the pain phenomenon. In addition, we present the relevant datasets used for automatic pain assessment and refer to the modalities and hardware requirements for developing an automatic pain assessment system. Section 2 presents the methodological approach for conducting the review, including our main research questions and search strategy. In Section 3, we describe the identified automatic pain estimation approaches. Section 4 includes the review's findings, and Section 5 concludes the review.

## 2. Review methodology & content

The present SLR has been conducted according to the “Guidelines for performing Systematic Literature Reviews in Software Engineering” by Kitchenham [37] and the PRISMA updated guideline for reporting systematic reviews [38].

### 2.1. Goal and research questions

This SLR aims to map the current research area of deep learning methods as applied to the automatic pain assessment domain. In particular, this study aims to address the following primary research questions:

1. What types of deep machine learning models are most commonly used, and what types of learning methods are used, *i.e.,* supervised, unsupervised, semi-supervised, self-supervised, *etc.,* for the automatic assessment of pain?

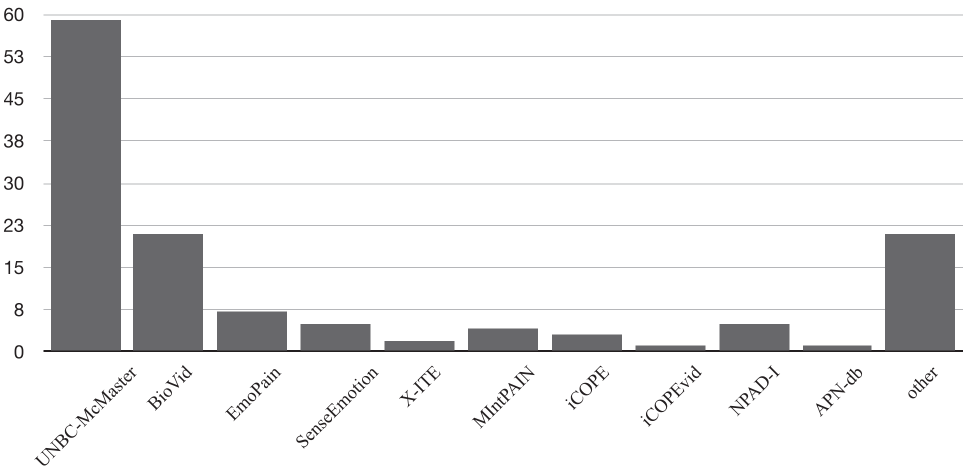

**Fig. 1.** The number of studies using the specific datasets. Several studies, however, employed more than one for contacting their experiments.

**Table 2**
Search Terms and Sources employed in the current review.

| Sources | Keywords |
|---|---|
| • Scopus[1]<br>• IEEE Xplore[2]<br>• ACM Digital Library[3] | • Pain Assessment<br>• Pain Detection<br>• Pain Recognition<br>• Pain Intensity<br>• Pain Estimation |
| **Search String:**<br>TITLE-ABS-KEY((“pain assessment”) OR (“pain detection”) OR (“pain recognition”) OR (“pain intensity”) OR (“pain estimation”)) AND (LIMIT-TO (DOCTYPE, “ar”) OR LIMIT-TO (DOCTYPE, “cp”)) AND (LIMIT-TO(LANGUAGE, “English”)) AND (LIMIT-TO (SUBJAREA, “COMP”)) | |

The specific search string employed to Scopus database. [1]https://www.scopus.com [2]https://ieeexplore.ieee.org [3]https://dl.acm.org

2. What modalities are used for the automatic assessment of pain, what are the most effective combinations of modalities, and what are the performance gains observed, if any, compared to unimodal approaches?

In parallel, we would also like to shed some light on how automatic pain assessment is most commonly approached, *i.e.*, a binary or multi-class classification problem, and identify if any benefits are reported when exploiting the temporal dimension of pain.

### 2.2. Search strategy

In order to conduct the review, we identified original studies searching digital libraries, utilizing keywords combined with boolean operators *OR, AND*, with no time constraints. The selected libraries were Scopus, IEEE Xplore, and ACM Digital Library since the major journals and conference proceedings related to affective computing are indexed in them. Further, the PubMed database was searched without additional results not identified from the previous databases. Regarding the search keywords, several terms were used in addition to “pain assessment” to identify all relevant studies. We note that we did not use any keyword associated with deep learning, though numerous terms are related to the particular type of approach. Table 2 presents the digital sources, the keywords, and an example of a completed search string.

All retrieved articles were investigated, but only studies that satisfied the following inclusion criteria were included: (1) are written in the English language; (2) are published in a journal or conference proceedings; (3) are related to the computer science research area and (4) affective computing field; (5) the employed modalities are video, biosignals, audio, movement data or combinations of them; (6) there are no time constraints since we want to review all relevant papers, including the early research efforts. Articles irrelevant to the inclusion criteria were removed. Subsequently, the following exclusion criteria apply: (1) research of pain assessment on animals; (2) the employed modalities are medical images, *e.g.*, X-rays, CT, and MRI, since they are not associated with affective computing; (3) articles that describe only theoretical concepts without implementation; (4) pain assessment on virtual agents or medical training dolls; (5) secondary literature, *e.g.*, review articles. Table 3 presents the applied eligibility criteria, where according to them, we accepted or rejected studies for the complete reading procedure.

### 2.3. Selection process

Once primary studies were identified, based on the previously described process, we proceeded with the following steps:

1. Export the citation for each paper, including the title, abstract, and keywords from each digital library.
2. Screening of the papers for verification and confirmation that meet the eligibility criteria.
3. In agreed papers that meet the eligibility criteria, a study of the entire paper was performed to confirm if the authors used deep learning approaches.

In total, 822 articles were retrieved from the databases employing the search keywords. After removing duplicates and those not fulfilling the eligibility criteria, 100 articles remained. In addition, the references of the papers selected were reviewed in order to potentially identify additional relevant studies. A list of 110 papers was included in the SLR, from which we extracted several metadata presented in Table 4, which assisted in obtaining insight and understanding about the approaches. In Fig. 2, we depict an overview of the SLR process based on [39]. In Fig. 3, we represent the whole procedure of the identification, screening, eligibility assessment, and inclusion for the final list of papers. Fig. 4 presents the number of studies per year and the machine learning methods used, *i.e.*, traditional feature engineering techniques or deep learning methods. The rapid increase of relevant pain assessment studies during the last ten years and the increased use of deep learning approaches since 2015 are evident.

## 3. Search results analysis about automatic pain assessment approaches

As already said, 110 scientific papers reporting automated pain assessment methods fulfilled the inclusion criteria defined, *i.e.*,

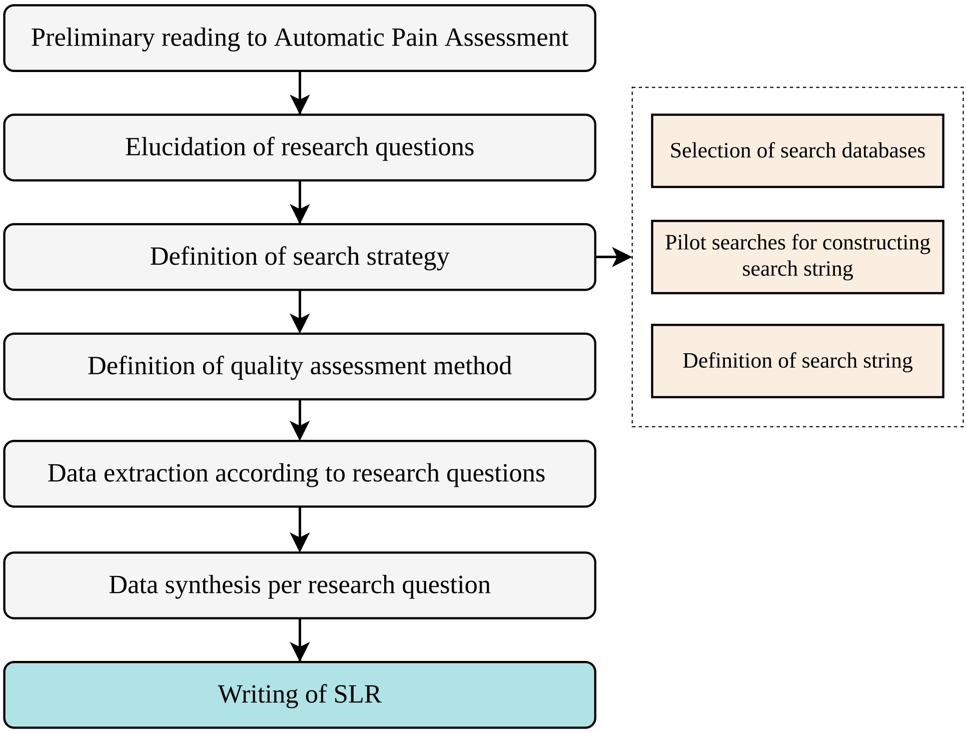
Preliminary reading to Automatic Pain Assessment
Elucidation of research questions
Definition of search strategy
Definition of quality assessment method
Data extraction according to research questions
Data synthesis per research question
Writing of SLR
Selection of search databases
Pilot searches for constructing search string
Definition of search string

**Fig. 2.** Overview of the systematic review.

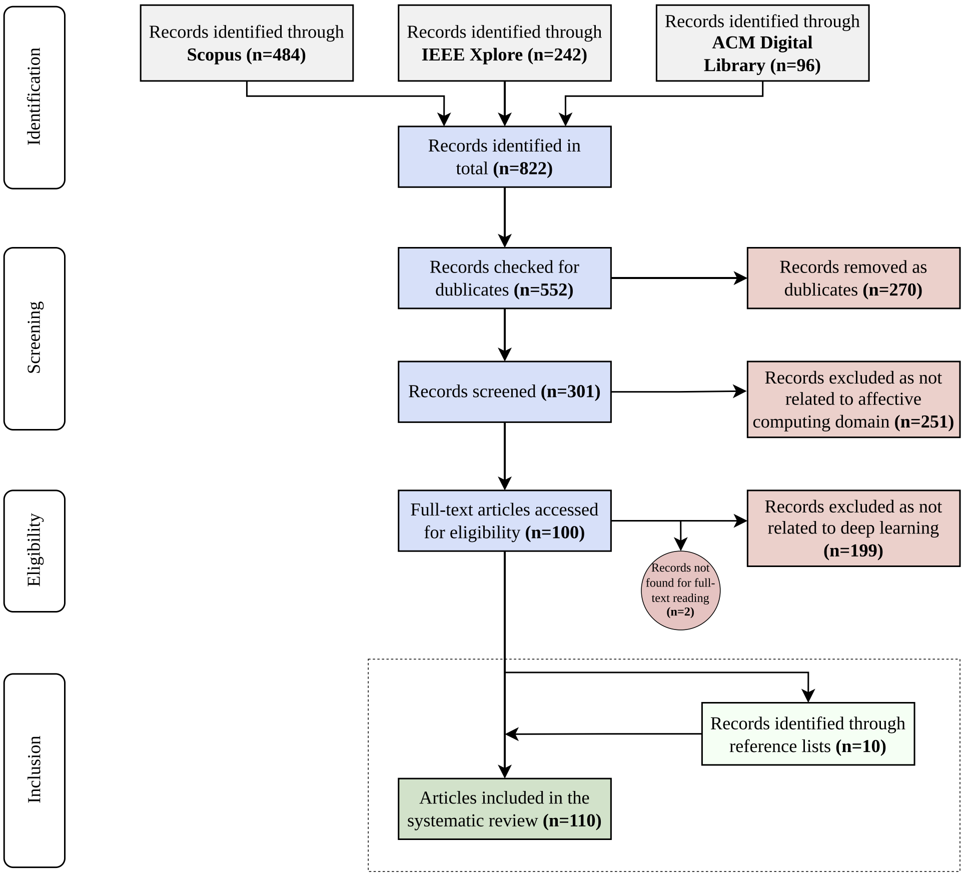
Identification
Screening
Eligibility
Inclusion
Records identified through Scopus (n=484)
Records identified through IEEE Xplore (n=242)
Records identified through ACM Digital Library (n=96)
Records identified in total (n=822)
Records checked for dublicates (n=552)
Records removed as dublicates (n=270)
Records screened (n=301)
Records excluded as not related to affective computing domain (n=251)
Full-text articles accessed for eligibility (n=100)
Records excluded as not related to deep learning (n=199)
Records not found for full-text reading (n=2)
Records identified through reference lists (n=10)
Articles included in the systematic review (n=110)

**Fig. 3.** Information flow of the systematic review.

**Table 3**
Eligibility Criteria applied on the studies.

| Eligibility Criteria | |
|---|---|
| Inclusion | Exclusion |
| • Language: English<br>• Type of papers: Journal articles, Conference proceedings<br>• Subject area: Computer Science<br>• Research field: Affective Computing<br>• Modalities: video, biosignals, audio, movement data<br>• No time constraints | • Animals as subjects of the research<br>• Medical images as information *e.g.*, X-rays, CT, MRI<br>• Papers where describe only theoretical concepts without implementation<br>• Use of virtual agents or medical training dolls as subject<br>• Review papers |

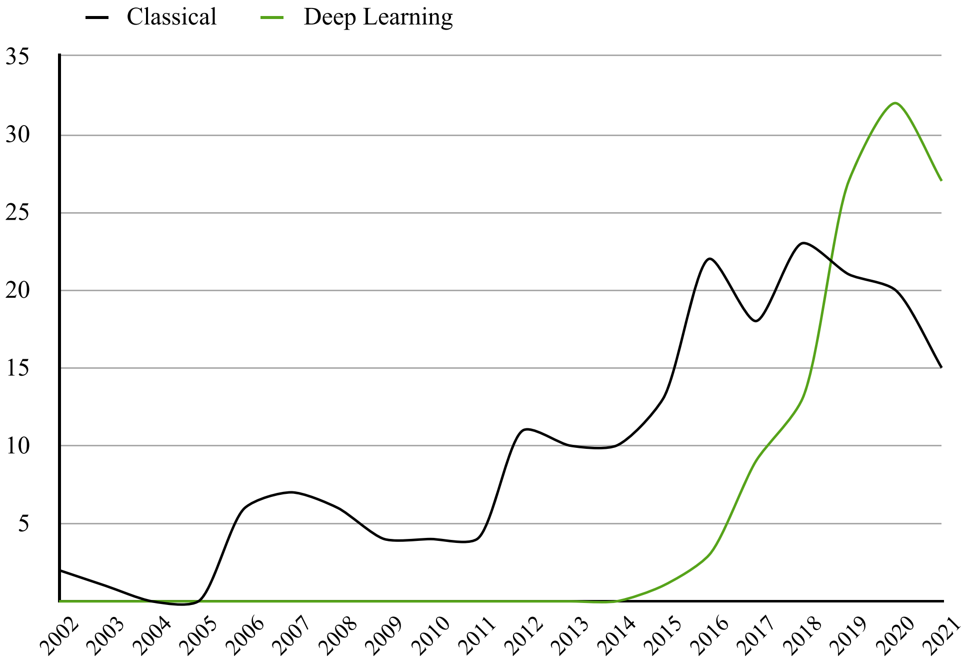

**Fig. 4.** The number of studies in automatic pain assessment by year of publication. The black line shows the studies based on classical machine learning and image/signal processing, while the green line shows the studies based on deep learning methods. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 4**
Extracted data from the included studies.

| Extracted Data | |
|---|---|
| • Year | • Learning Method |
| • Title | • Pre-trained Model |
| • Unimodal, Multimodal | • Classification, Regression |
| • Modality | • Objective Ground Truth |
| • Temporal Exploitation | • Interpretation |
| • Fusion Method | • Validation Method |
| • Deep Model | • Number of Subjects |
| • Non-Deep model | • Performance Metrics |
| • Non-Deep learned features | • Dataset |

they report application of deep learning methods and are included in our analysis. The initial separation of these papers relates to the number of information channels used, *i.e.*, unimodal vs. multimodal. Subsequently, the unimodal approaches are distinguished as 1) vision-based, 2) contact-sensor-based, and 3) audio-based. It is worth pointing out that the highest percentage of the studies report vision-based approaches. Specifically, 84 of the 110 papers exploit the face as an input channel.

Additionally, a further subdivision was done to vision-based approaches regarding the temporal exploitation of the modalities. The studies which utilized the temporal dimension are divided into 1) non-machine learning-based and 2) machine learning-based approaches, which in turn are divided into explicit and implicit machine learning-based approaches. The non-machine learning methods are based on the dynamic encoding of the initial information, *e.g.*, optical flow features or the subtraction of subsequent frames. Machine learning-based methods are based on a learning procedure; the explicit approaches develop specific/individual temporal modules, *e.g.*, a long short-term memory network (LSTM). The implicit approaches focus on extracting temporal features from the models' training process, *e.g.*, 3D convolutional neural network (3D CNN).

It should be pointed out that the direct comparison of the performances reported in the various studies is not always possible since, in many cases, different scales and ground truths were used. In addition, the studies differ even when using the same dataset. For example, some studies may exclude specific subjects from the experiments without reporting this, thus making comparability impossible, an issue also reported in [18]. For these reasons, we note that proper comparisons can be made between studies if the following criteria are fulfilled: employ the same dataset or part of it, adopt the same validation method, implement an identical task (*e.g.*, pain detection, multi-level pain estimation), and the same performance metrics are used. Furthermore, in this SLR, we do not scrutinize the processing techniques such as face detection, alignment, resizing, *etc.*, since our emphasis is on the learning procedure of pain features.

### *3.1. Unimodal approaches*

In this section, we present the studies that utilized only one information channel to estimate the subject's pain condition.

#### 3.1.1. Vision-Based: Non-temporal exploitation

The first publicly available pain database which contributed significantly to the progress and development of automatic pain assessment methods was the UNBC-McMaster Shoulder Pain. A plethora of studies has employed the particular dataset. Pedersen [40] in 2015 implemented the first deep learning approach to address the pain assessment problem, utilizing a 4-layer contractive autoencoder. He exploited the encoding representation along with a support vector machine (SVM) and achieved high performance of pain detection at the frame level. An important contribution to the vision-based approaches for pain recognition has been the EmoPain challenge in 2020. It was the first international competition focused on creating a platform for comparing machine learning methods of automatic chronic pain assessment. Egede et al. [41] presented the *EMOPAIN 2020 Challenge*, in which the corresponding dataset consists of extracted features using handcrafted approaches and deep learned models. The authors utilized facial landmarks, histogram of oriented gradients (HOG), and deep vectors elicited from *VGG-16* [42], and *ResNet-50* [43] accordingly, which are pre-trained on the *Aff-Wild* dataset[1]. The authors report that the hand-engineered features combined with deep learning cues obtained the highest performance. Likewise, Yang et al. [44] utilized low and high-level cues extracted from local descriptors and pre-trained *VGG-16* CNN [42], respectively, and combined them employing weighted coefficients. Semwal and Londhe [45] showed that the fusion of deep-learned features with facial landmarks is beneficial for multi-class pain estimation. Lakshminarayan et al. [46] exploited deep learned and handcrafted features, *i.e.,* learned features from *VGG-16* [42] and *ResNet-50* [43], HOG, action units occurrence, action units intensity, facial landmarks, and head pose through a fully connected network. The study revealed that combining *VGG-16* and handcrafted features led to lower regression error. However, in [47], the authors achieved maximal performance utilizing the *VGG-16* features exclusively with a similar fully connected network as a classifier.

On the contrary, Semwal and Londhe [48] report that the conventional handcrafted feature engineering method has several drawbacks and difficulties in its application, while deep neural networks are highly computationally expensive. Therefore, they suggest the deployment of a relatively shallow 4-layer CNN, in which the computational training cost is reduced because of the limited number of parameters. However, the performance and outcomes are comparable to those obtained using deeper architectures. A different approach emanated from [49], in which the authors focused on representing the facial expressions as a compact binary code for the classification of different pain intensity levels, with a pre-trained model [50] conducting feature extraction and a fully connected network constructing the binary code.

Other approaches employed CNNs ensemble designs with different architectures to exploit variations of characteristics. Semwal and Londhe [51] utilized three compact CNNs, *VGG-16* [42], *M-MobileNet* [52], and *GoogleNet* [53], integrating their predictions using the average ensemble rule. The experiments demonstrate that merging the CNNs leads to better classification performance than using them individually. Kharghanian et al. [54] reported the development of a convolutional deep belief network (CDBN) through an unsupervised feature learning approach. The extracted features were used by an SVM in order to distinguish two states for the binary classification of pain, *i.e.,* pain and no pain. In subsequent research, [55], the authors added two additional layers to the CDBN. Unfortunately, the results are not comparable since the evaluation methods were different.

Several papers claim that since the presence and manifestation of pain are visible in certain areas of the face, it would be advantageous to exploit these areas as input information instead of the entire facial image. This way, the model will be trained to pay attention to the regions most relevant to the manifestation of pain, excluding noise. Along these lines, Huang et al. [56] initially detected the face regions of the left eye, right eye, nose, and mouth, followed by a multi-stream CNN responsible for the feature extraction, consisting of 4 sub-CNNs, one for each region. A notable element of the particular framework was that the extracted features were assigned with learned weights to provide the required attention to account for the fact that each region contributes differently to pain expression. Similarly, in [57], a 9-layer CNN was combined with an attention mechanism assigning different weights related to the expressiveness of the face regions to generate attention face maps which provided up to 19% more accurate predictions. In [58], the authors proposed a multi-scale regional attention network (MSRAN), which utilizes several cropping regions of each video frame. Furthermore, it incorporates a self-attention and a relation attention module to emphasize pain-related regions and explore their relationship. Beyond the creation of saliency maps, the work reported by Li et al. [59] relied on the idea of [60], combining contrastive and multi-task training through an autoencoder. In addition, similar to basic facial expression recognition, one challenge for pain intensity estimation is that some individual characteristics, *e.g.,* face shapes, may cause great diversities in the same emotion. As a result, it is usually challenging to distinguish two adjacent intensity levels of pain expression as each intensity has a significant variation. In addressing this issue, Peng et al. [61] scrutinized facial shape information and developed a deep multi-task network that interpreted the relationship between pain recognition and shape, which indeed enhanced the performance of pain estimation. Similarly, Xin et al. [62] presented a novel multi-task framework incorporating a CNN feature learning module combined with an autoencoder attention component, estimating the subject identity as well since different individuals have specific pain manifestations. The experiments showed state-of-the-art performance in publicly available datasets.

Most research efforts report the results of experiments conducted in controlled lab settings, with proper lighting, low variability of head pose, and absence of occlusions. These are not representative of the typical hospital environments. The authors of [63] focused on pain assessment to study the problem of pain assessment in uncontrolled environments. Developing a relatively shallow CNN with 3 convolutional layers achieved high multi-class classification performance comparable to deeper pre-trained models. In a follow-up study by the same authors [64], a more complex deep framework was developed, consisting of 3 modules. Exploiting high-level spatial feature descriptors with local and global geometric cues, they achieved results comparable to those obtained from other models, such as *GoogleNet* [65] and *VGG* [42]. Lee and Wang [66] explored the intensive care unit (ICU) setting, where the phenomenon of partially occluded faces frequently occurs, which creates difficulties in every facial analysis task. Concerning the feature extraction method, they developed a 4-layer CNN combined with an extreme learning machine (ELM) network for the final estimation. In the work reported in [67], the authors have used CNNs to implement a novel approach to evaluating the timing information of pain by classifying the section of frames where the pain was triggered, reached its climax, and started to diminish. Nugroho et al. [68] studied the issue of pain detection in the context of a smart home-care setting, in which small and relatively computationally weak mobile devices are used for detecting and classifying individuals' pain, particularly elders. Utilizing and modifying *Open-*

[1] https://ibug.doc.ic.ac.uk/resources/first-affect-wild-challenge

*Face*[2], a face recognition library based on the pre-trained *FaceNets* [69], revealed that with transfer learning, the binary classification of pain (*i.e.,* pain vs. no pain) is possible to be estimated in real-time, even in low-powered hardware.

Several researchers [70] [71] point out that in the majority of studies, the models, either with deep or shallow architectures, are trained on dataset-specific features and not on actual pain-related features. Also, most studies investigating the facial expression of pain implement validation methods on the same database, whereas cross-database performance is less considered. These result in approaches that are not applicable in real-world settings. In addressing these issues, Dai et al. [70] experimented with the combination of pain and emotion-detection datasets in order to develop a real-time pain assessment system with higher generalization capabilities. Their work highlights the importance of cross-corpus evaluation, real-time testing and the need for a well-balanced and ecologically valid [72] pain dataset.

A number of papers exploited combinations of different pain scales as ground truth in order for the final prediction to be as objective and reliable as possible. Liu et al. [73] focused on the estimation of individual's pain via a two-stage personalized model, trained with active appearance model (AAM) facial landmarks in a multi-task manner, and VAS and observed pain index (OPI) scores as ground truth. In a similar manner, Xu et al. [74] combined the ground truth utilizing various pain scales simultaneously in order to reduce the mean square error (MSE) using the *VGG-Face* [75]. On the other hand, the authors in [76] report that the original ground truth has limitations related to the subjects themselves or data annotation experts. For this reason, they re-annotated the dataset employing multi-expert judgments via seven evaluators. Based on the multidimensional scaling method, they mapped the frames to illumination-invariant 3D space to feed a pre-trained *AlexNet* [77].

Celona and Manoni [78] explored the facial expression of neonates to detect the presence or absence of pain. The authors experimented with hand-crafted features but achieved the highest performance when utilizing two pre-trained models, namely *VGG-Face* [75] and mapped LBP+CNN (MBPCNN) [79]. In parallel, the authors of [80] also report that the pre-trained models are essential since training from scratch with small datasets, such as neonatal ones, will cause over-fitting. They achieved their highest classification scores by employing and fine-tuning the *VGG-16* [43] as a whole, instead of only its last layers. Contrary to the findings of various studies indicating that using pre-trained models was the most beneficial approach when applied to neonates, Zamzmi et al. [81] report a model's design and training process from scratch in an end-to-end manner. They mention that most face recognition methods and techniques are designed for adults who have different facial structures and diverse ways of pain manifestations than infants. By introducing a lightweight 2D CNN, they achieved high performance in pain detection. However, its external validation on a different neonatal dataset revealed challenges regarding its generalizability. In 2019, Brahnam et al. [31] presented a new neonatal video dataset called *iCOPEvid*. This was a significant contribution since, until then, the only publicly available neonatal pain dataset [30] included only static images. Using local descriptors based on bag-of-features (BoF) outperformed the deep learning-based results obtained with *VGG-Face* [75] and *ResNet* [43]. In addition, the combination of hand-crafted and deep-learned features has a negligible increase in the system's performance. Contrary to these findings, the authors in [82] report that for the binary classification (*i.e.,* pain vs. no pain) problem, the most successful approach was based on the fusion of higher-level features of a *VGG* [83] and optical flow strains through a naive Bayes as a classifier. In another study [84], the authors implemented a Wasserstein generative adversarial network with gradient penalty (WGAN-GP) [85], demonstrating that the augmentation of the training set by the generated synthetic samples improved the classification performance. Table 5 summarizes the vision-based studies, which are based solely on the spatial dimension.

### *3.1.2. Vision-Based: Exploitation of the temporal dimension (Non-machine learning-based)*

The complex and dynamic nature of pain makes its assessment very challenging. The use of static and independent visual frames is incapable of capturing the temporal evolution of the phenomenon and thus often leads to erroneous pain estimation. In addition, several studies point out that applying deep learning methods to limited-size datasets is problematic, and a proposed solution combines deep learning with traditional feature extraction techniques. As a result, Egede et al. [86] elicited deep features from a pre-trained CNN corresponding to eyes and mouth regions. They employed a relevance vector regressor (RVR), demonstrating that the combined exploitation of deep and hand-crafted features achieves the highest performance. Although the UNBC-McMaster database provides valuable pain information material, it is characterized by imbalanced samples (*i.e.,* limited number of frames manifesting pain), challenging deep learning researchers. Egede and Valstar [87], when facing the specific problem, developed a method that capitalized on the observation that neighboring classes of pain levels share a large number of common characteristics. This led them to a decision that for classes with a limited number of samples, it is not necessary to extract every possible type of feature because certain features have already been exploited from other disparate classes. In addition, the study reported that the combination of hand-crafted and deep-learned features achieved improved performance. In contrast, in a subsequent study [88] by the same authors, the identical approach was employed to minimize data imbalance. However, they extracted deep-learned features exclusively, and as reported, they could not achieve equivalent performance levels.

Tavakolian et al. [89] studied the phenomenon of pain from a different perspective. The specific research objective was to detect genuine versus acted pain based on its facial manifestation, which is valuable for medical and criminal applications. The authors designed a residual GAN (R-GAN) exploiting the subtle facial changes and also captured the dynamic nature of facial expressions utilizing a weighted spatio-temporal pooling method (WSP). In a subsequent study [90], the authors suggest that self-supervised learning is recommended to reduce time and effort in collecting labeled data since this approach does not require annotation of the entire dataset. The authors introduce a novel similarity function to learn generalized representations using a Siamese network. They employ statistical spatio-temporal distillation (SSD) based on the Gaussian scale mixture (GSM) to make the method computationally efficient. In this way, they encode the spatiotemporal variations of the facial video into a single RGB image and avoid more complex models.

Other authors also attempt to capture the dynamic nature of pain. For example, in [91] the authors combined a random forest classifier and the pre-trained *MobileNetV2* model [92], encoding each video into an image by selecting and merging three frames from different time steps. Othman et al. [93] report that to achieve better estimation results and to improve the model's generalizability, it is important to use datasets that include diverse distributions of age and gender and even various poses, occlusions, lighting conditions, *etc.* The authors deployed numerous data combinations utilizing a reduced version of *MobileNetV2* [92] and pointed out that cross-data training is valuable, respecting the generalization affair.

[2] http://cmusatyalab.github.io/openface

**Table 5**
Studies utilized camera-based information with non-temporal exploitation.

| Paper | Input | | | Processing | | | | Evaluation | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Modality | Non deep features | Fusion M/E | Deep model | Non deep model | Learning Method | Classific. /Regres. | Objective | GT | Number subjects | Validation Method | Dataset | Metrics |
| '19 [31] | F (RGB) | texture descriptors | - FF | 2D CNN$^{+}$ | SVM | SL | C | P | O | 49 | k-fold | iCOPEvid | 79.80 AUC |
| '15 [40] | F (RGB) | - | - - | AE | SVM | SeSL, SL | C | P | PS | 25 | LOSO | UNBC | 86.10 ACC, 96.50 AUC |
| '20 [41] | F (RGB) | - | - FF | 2D CNN$^{+}$ | NN | SL | R | IC | O | 36 | hold-out | EmoPain | 0.91 MAE$^{\ddagger}$ |
| '18 [44] | F (RGB) | HOG, statistics | - FF | 2D CNN$^{+}$ | SVR | SL | R | IC | PS | 25 | LOSO | UNBC | 1.44 MSE$^{\ddagger}$ |
| '21 [51] | F (RGB) | - | - DF | 2D CNN$^{+}$ | - | SL | C | ID | PS | 25 | k-fold | UNBC | 93.87 ACC$^{\ddagger}$ |
| '16 [54] | F (RGB) | - | - - | CDBN | SVM | UL | C | P | PS | 25 | LOSO$^{\dagger}$ | UNBC | 87.20 ACC$^{\ddagger}$ |
| '21 [55] | F (RGB) | - | - - | CDBN | SVM | SL | C | P | PS | 25 | LOSO | UNBC | 93.16 AUC |
| '19 [56] | F (RGB) | - | - FF | 2D CNN | - | SL | C | ID$^{1}$, IC | PS | 25 | LOSO | UNBC | 88.19$^{1}$ ACC |
| '20 [57] | F (RGB) | - | - - | 2D CNN | - | SL | C | ID | PS | 25 | hold-out | UNBC | 51.10 ACC$^{\ddagger}$ |
| '20 [61] | F (RGB) | - | - FF | 2D CNN$^{+}$ | - | SL | R | ID | S | 25 | ? | UNBC | 79.94 ACC$^{\ddagger}$ |
| '21 [63] | F (RGB) | - | - - | 2D CNN | - | SL | C | ID | O | 8 | k-fold | other | 97.48 ACC$^{\ddagger}$ |
| '19 [66] | F (RGB) | - | - - | 2D CNN | ELM | SL | R | IC | PS | 25 | k-fold | UNBC$^{\bullet}$ | 1.22 MSE$^{\ddagger}$ |
| '19 [67] | F (RGB) | - | - - | 2D CNN | - | SL | C | TR, CL, DI | PS | 25 | k-fold | UNBC | 60.00 ACC |
| '19 [71] | F (RGB) | - | - - | 2D CNN$^{+}$ | - | SL | C | AUs-D | PS | 25, 43 | k-fold | UNBC & CK+$^{1}$, Wilkie | 97.70$^{1}$ ACC$^{\ddagger}$ |
| '17 [73] | F (RGB) | statistics | - - | NN | GPM | WSL | R | IC | O, S | 25 | k-fold | UNBC | 2.18 MAE |
| '20 [74] | F (RGB) | statistics | - FF | 2D CNN$^{+}$ | NN | SL | R | IC | S | 25 | k-fold | UNBC | 1.95 MAE$^{\ddagger}$ |
| '19 [76] | F (RGB) | LBP, MDS | - - | 2D CNN$^{+}$ | - | SL | C | ID | O | 25 | hold-out | UNBC | 80.00 ACC |
| '18 [80] | F (RGB) | - | - - | 2D CNN$^{+}$ | - | SL | C | ID | O | ? | hold-out | other | 78.30 ACC |
| '18 [82] | F (RGB) | optical flow | - FF | 2D CNN$^{+}$ | SVM, kNN, NB | SL | C | P | O | 31 | k-fold | other | 92.71 ACC, 94.80 AUR |
| '19 [84] | F (RGB) | - | - - | WGAN-GP | - | SL | C | P | O | 26 | LOSO | iCOPE | 93.38 ACC |
| '17 [107] | F (RGB) | - | - - | 2D CNN$^{+}$ | - | SL | R | IC | PS | 25 | LOSO | UNBC | 0.99 MAE$^{\ddagger}$ |
| '20 [108] | F (RGB) | - | - - | 2D CNN | - | SL | C | ID | ST | 87 | hold-out | BioVid (A) | 36.60 ACC |
| '20 [109] | F (RGB) | - | - - | 2D CNN | - | SL | C | P | PS | 25 | hold-out | UNBC | 97.00 PPV$^{\ddagger}$ |

$^{+}$: Pre-trained model -:Not exist &: in Dataset indicates the utilization of cross-database training/validation ?: Not found $^{\dagger}$: The authors provide additional experiments with other validation methods $^{\bullet}$: The authors utilized occluded facial images $^{\ddagger}$: The authors provide additional metrics **Modality:** F: face region **Non deep features:** LBP: local binary pattern MDS: multidimensional scaling **Fusion:** M: fusion of modalities E: fusion of deep learned features or hand-crafted features **Deep models:** AE: autoencoder RCNN: recurrent convolutional neural network CDBN: convolutional deep belief network CNN: convolutional neural network NN: neural network WGAN-GP: Wasserstein generative adversarial model with gradient penalty **Non deep model:** SVM: support vector machine GPM: Gaussian process regression model kNN: k-nearest neighbors NB: naive Bayes ELM: extreme learning machine **Learning Method:** SL: supervised learning SeSL: semi-supervised learning UL: unsupervised learning WSL: weakly supervised learning **Classific./Regres.:** C: classification R: regression **Objective:** P: presence of pain ID: intensity in discrete scale IC: intensity in continuous scale TR: trigger CL: climax DI: diminishing AUs-D: Action Units detection **GT**: ground truth PS: Prkachin and Solomon S: self-report O: observer rating ST: stimulus **Validation Method:** LOSO: leave one subject out **Metrics:** AUC: Area Under the ROC Curve ACC: accuracy PPV: precision MSE: mean squared error MAE: mean absolute error

**Table 6**
Studies utilized vision-based information with non-temporal exploitation.

| Paper | Input | | | Processing | | | | Evaluation | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Modality | Non deep features | Fusion M/E | Deep model | Non deep model | Learning Method | Classific. /Regres. | Objective | GT | Number subjects | Validation Method | Dataset | Metrics |
| '21 [45] | F (RGB) | facial landmarks | - FF | 2D CNN | NN | SL | C, R | ID, IC$^{1}$ | P | 25 | LOSO$^{\dagger}$ | UNBC | 0.17$^{1}$ MSE$^{\ddagger}$ |
| '20 [46] | F (RGB) | HOG, head pose, AUs intensity/ occurrence, facial landmarks, | FF - | 2D CNN$^{+}$ | NN | SL | R | IC | O | 36 | hold-out | EmoPain | 5.48 RMSE$^{\ddagger}$ |
| '20 [47] | F (RGB) | - | - - | 2D CNN$^{+}$ | NN | SL | R | IC | O | 36 | hold-out | EmoPain | 1.49 RMSE$^{\ddagger}$ |
| '18 [48] | F (RGB) | - | - - | 2D CNN | - | SL | C | ID | PS | 25 | hold-out | UNBC | 92.00 ACC$^{\ddagger}$ |
| '18 [49] | F (RGB) | statistics, distance metrics | - FF | 2D CNN$^{+}$ | - | SL | C, R | ID, IC | PS | 25 | LOSO | UNBC | 0.81 PCC 0.69 MSE |
| '21 [58] | F (RGB) | - | - FF | 2D CNN$^{+}$ | - | SL | C, R | ID, IC | P | 25 | LOSO | UNBC | 91.13 ACC, 0.78 PCC, 0.46 MSE |
| '18 [59] | F (RGB) | - | - - | AE$^{+}$ | - | SL | R | IC | PS | 25 | k-fold | UNBC | 0.33 MAE$^{\ddagger}$ |
| '21 [62] | F (RGB) | - | - FF | [AE, 2D CNN]$^{\cup}$ | - | SL | C, R | ID$^{1}$, IC$^{2}$, P$^{3}$ | P, ST | 25, 87 | LOSO | UNBC$^{1}$, BioVid (A)$^{2}$ | 89.17$^{11}$ ACC, 0.81$^{21}$ PCC, 85.65$^{32}$ ACC, 40.40$^{12}$ ACC |
| '21 [64] | F (RGB) | entropy texture descriptors | - - | 2D CNN$^{+}$ | - | SL | C | ID | O | 8 | k-fold | other | 0.92 PPV$^{\ddagger}$ |
| '18 [68] | F (RGB) | - | - - | 2D CNN$^{+}$ | - | SL | C | P | PS | 14 | k-fold | UNBC | 93.00 ACC |
| '19 [70] | F (RGB) | - | - - | 2D CNN | - | SL | C | P | PS | 25, 20 | k-fold | UNBC & BioVid (A)$^{\diamond}$ | 56.75 ACC |
| '17 [78] | F (RGB) | HOG, LBP | - FF | 2D CNN$^{+}$ | SVM | SL | C | P | O | 26 | LOSO | iCOPE | 73.78 ACC |
| '19 [81] | F (RGB) | - | - - | 2D CNN | - | SL | C | P | O | 31 | LOSO | NPAD$^{1}$, iCOPE$^{2}$ | 96.98$^{1}$ ACC$^{\ddagger}$, 89.80$^{2}$ ACC |
| '21 [110] | F (RGB) | - | - - | 2D CNN$^{+}$ | - | FL | C | P | PS | 25 | LOSO | UNBC | 76.00 ACC$^{\ddagger}$ |
| '21 [111] | F (RGB) | - | - - | 2D CNN$^{+}$ | - | SL | C | P | O | 25 | hold-out | UNBC | 75.49 ACC |
| '21 [112] | F (RGB) | - | - - | 2D CNN$^{+}$ | SVR | SL | R | IC | P | 25 | LOSO | UNBC | 0.34 MSE |
| '21 [113] | F (RGB) | - | - - | 2D R-CNN | - | SL | C | P | O | ? | hold-out | other | 87.80 PPV |
| '21 [114] | F (RGB) | - | - - | 2D CNN | - | SL | C | ID | P | 28 | LOSO$^{\dagger}$ | UNBC | 90.30 ACC |
| '19 [115] | F (RGB) | - | - - | 2D CNN | - | SL | C | P | O | 31 | hold-out | NPAD$^{1}$, iCOPE$^{2}$ | 91.00$^{1}$ ACC$^{\ddagger}$, 84.50$^{2}$ ACC$^{\ddagger}$ |
| '21 [116] | F (RGB) | - | - - | 2D CNN$^{+}$ | - | SL | C | P | O | 26, 30 | hold-out | iCOPE & UNIFESP | 89.90 ACC$^{\ddagger}$ |
| '21 [117] | F (RGB) | - | - - | 2D CNN | - | SL | C | AUs-D | P | 10 | hold-out | Pain-ICU | 77.00 ACC$^{\ddagger}$ |

$^{\cup}$: The authors combined the deep models into a unified framework $^{\diamond}$: The authors experimented with additional datasets combinations **Non deep features:** AUs: actions units HOG: histogram of oriented gradients **Non deep model:** SVR: support vector regression **Learning Method:** FL: federated learning **Metrics:** RMSE: root mean squared error

### 3.1.3. Vision-Based: Exploitation of the temporal dimension (Implicit)

Considering the application of 3D CNNs, several studies employ this particular approach. Specifically, Tavakolian and Hadid [94] developed a 3D CNN capturing the dynamic facial representation from videos. The authors report that usually, the researchers using 3D convolution techniques deploy a fixed temporal kernel depth. This results in the ineffective implementation of simultaneously extracting short, mid, and long ranges from the sequences. They designed a model having parallel 3D convolutional layers of variable temporal depths capable of capturing temporal dependencies from 32 consecutive frames as a time window. Similarly, Wang and Sun [95] exploited 3D convolutions based on the architecture reported in [96], which consisted of 8 convolutional layers with 3x3x3 filters. Although a very high performance is reported, the authors also state that the deep features are inefficient, if extracted from small databases, since the models have difficulty generalizing appropriately. Interestingly, in [97], the authors have created a framework that integrated 3D, 2D, and 1D CNNs, which are used for extracting spatio-temporal, spatial, and geometric features correspondingly. In relation to 3D CNN, they have modified the architecture reported in [98], by combining discrete kernels of 1x3x3 and 3x1x1 rather than the classic 3D kernel of 3x3x3. Other authors have also developed a 3D deep CNN with various temporal depths, based on the assumption that using 3D kernels with different ranges it will be feasible to capture short, mid, and long-range facial expression variations [99]. Further, taking into consideration that the training process of a deep 3D CNN from scratch is difficult and time-consuming, they introduced a cross-architecture knowledge transfer learning technique, which utilizes a pre-trained 2D CNN in the training of the 3D CNN. In the work reported in [100] and [101], the authors have adopted the weak-supervised domain adaptation, in which the source domain was related to human affective expressions, and the target domain were data based on pain expressions explicitly. Their proposed framework included an inflated 3D-CNN (I3D) [102] designed with 3 convolutional layers and 3 inception modules [53] to exploit both spatial and temporal information from the videos.

Bargshady et al. [103] exploited the HSV instead of the RGB color space, since they advocated that it is more valuable for tasks related to human visual perception, *e.g.,* skin pixel detection and multi-face detection. The authors utilized the pre-trained *VGG-Face* [75] for feature extraction. This was followed by a temporal convolutional network (TCN) based on dilated causal convolutional operations exploiting the temporal dependencies. Rezaei et al. [104] tried to tackle the problem of pain detection in people with dementia, which is incredibly challenging because the existing pain datasets do not include an adequate amount of images or videos of such (elderly) subjects. They designed a 2D CNN of 10 layers receiving pairs of pain and no-pain images, analyzing changes from frame to frame, and training in a multi-task manner, utilizing the contrastive training method [105]. The authors report high-performance rates both in healthy and in people with dementia. The authors in [106] studied the potential use of the shallowest-possible 1D CNN architectures for pain recognition in real-time settings, extracting facial action units from each frame with the *OpenFace 2.0*[3] toolkit with promising results.

### 3.1.4. Vision-Based: Temporal exploitation (explicit)

Several efforts have focused on alleviating the limitations of using static frames and developing dedicated temporal modules. Zhou et al. [118] addressed the specific problem by deploying a regression framework based on a 4-layer recurrent convolutional neural network (RCNN), each with a length of 3 time steps. Rodriguez et al. [119] exploited the dynamic information, designing an LSTM, and fed it with the extracted vectors from the *VGG-16* [43]. Similarly, the authors in [120] stated that since facial expressions change over time, it is necessary to study the spatio-temporal dimension of pain. An improved estimation performance has been achieved through a fine-tuned 16-layer CNN model [75], an LSTM using 16 frames as a time window, and super-resolution techniques. By performing a combination of the CNN *VGG-Face* [75] and a 3-layer LSTM the authors in [121] extracted spatio-temporal features from grayscale images in which they applied zero-phase component analysis (ZCA), while in [122] principal component analysis (PCA) was adopted to reduce the dimensionality. Similarly, Mauricio et al. [123] deployed the *VGG-Face* [75], but instead of LSTM they utilized a 2-layer gated recurrent unit (GRU). The authors in [124] utilized a conventional 2D CNN and two RCNNs extracting temporal features, exploiting both previous and subsequent frames, to exploit the time dimension of the expressions.

In a subsequent study [125], a similar approach was followed regarding the feature extraction from *VGG-Face* [75]. However, the authors exploited ensemble learning to create three distinct modules of CNN-biLSTMs, whose outputs were merged to produce the final prediction. Salekin et al. [126] employed a bilinear CNN (B-CNN) based on the standard CNN architecture of *VGG* [42], with models pre-trained on the *VGGFace2*[4] and *ImageNet*[5] datasets, respectively. In addition, an LSTM model exploited the temporal dependencies from the image sequences. Kalischek et al. [127] studied the application of deep domain adaptation to facial expression and pain detection, employing the self-ensembling approach [128], in which the training process was evolving in an unsupervised manner with a long-term recurrent convolutional network (LRCN). Despite the fact that state-of-the-art performance was achieved with self-ensembling regarding facial expression recognition, the results were relatively poor regarding pain recognition. This may be a consequence of the subtle expressions in the pain events.

There is a limited number of studies regarding multi-task approaches, although the pain datasets provide the additional required information. Martinez et al. [129] introduced a novel personalized multi-task machine learning method based on individual physiological and behavioral pain response profiles for pain estimation. They initially extracted AAM facial landmarks and drove them to a bidirectional LSTM (biLSTM), producing PSPI scores to predict the final VAS. At the same time, the authors in [130] utilized the *AlexNet* [77] with 2 GRU layers to capture the temporal dependencies, combined with self and observer-reported pain intensity as ground truth. Vu et al. [131] also developed a multi-task framework to estimate the pain level and simultaneously reconstruct heatmaps from the predefined action unit locations. In this way, the model generalized better, while a CNN combined with an LSTM exploited the micromovements between the frames.

In [132], the authors noted that beyond the pain manifestation on specific facial areas, certain frames exhibit pain expressions more vividly in a video sequence, requiring appropriate handling. For this reason, they developed a novel framework with attention saliency maps through a *VGG-16* [42], GRUs, and learned weights associated with the contribution of each frame in the final pain intensity estimation. The study revealed that there are opportunities to achieve compelling performance through the exploitation of dynamic and salient features. At the same time, the authors in [133], through the *VGG-11 (configuration A)* [42] and an LSTM, they developed an attention mechanism of various convolutional filters predicting pain intensity from 16 consecutive frames.

[3] https://github.com/TadasBaltrusaitis/OpenFace

[4] https://www.robots.ox.ac.uk/~vgg/data/vgg_face

[5] https://www.image-net.org

Xu and Liu [134] interestingly utilized a *ResNet-50* [43] model with a learned attention mechanism extracting spatial features, followed by a transformer encoder [135] to exploit the sequential nature of video frames achieving promising results.

Other studies, such as [136], employed the extracted action units to train a 2-layer LSTM predicting an 11-point scale of pain, adopting curriculum learning. In [137], a convolutional LSTM (C-LSTM) network was developed to extract spatial and temporal features from videos simultaneously, demonstrating significant differences among temporal and non-temporal models and revealing the importance of the time dimension for the accurate pain estimation. Other authors, *e.g.,* Rasipuram et al. [138], utilized videos in the wild for the detection of pain. Their approach was based on a pre-trained network [139], which generated a 3D morphable model of the face without exploiting facial landmarks, and was combined with an LSTM. Zhi and Wan [140] introduced sparse coding with LSTM (SLTM) based on the iterative hard thresholding algorithm (ISTA) [141] to capture the dynamic nature of facial expressions. The authors utilized the SLTMs solely for spatial and temporal feature extraction, converting the frames to grayscale format and resizing them to 32x32 pixels. The proposed approach did not achieve high performance but may be convenient in cases where speed and efficiency are required. Finally, the authors in [142] employed implicit and explicit approaches to exploit the temporal aspect of image sequences. They created motion history and optical flow images from the original frames, accompanied by a 10-layer CNN combined with a 2-layer biLSTM. The study demonstrates that the weighted score aggregation of the motion history and optical flow provides improved performance. Table 7 summarizes all the studies that exploit the modalities' temporal dimension.

### 3.1.5. Contact sensor-Based

Contact sensors are an alternative option for assessing pain, and the pain estimation results are often superior to those based on vision. Table 10 presents the studies that report using the information from contact sensors to assess pain. Yu et al. [143] analyzed three classes of pain, namely no pain, moderate and severe, based on EEG signals. By extracting several bands from the biosignals (*i.e.,* alpha, beta, gamma) and employing a convolution module for each band, the authors report that the combination of bands resulted in improved results as compared to using them independently. Similarly, the authors of [144] employed EEG potentials and an autoencoder to encode the raw data into a compressed format, utilizing the logistic regressor as the classifier.

Other researchers, such as Rojas et al. [145], exploited functional near-infrared spectroscopy (fNIRS) to estimate the pain condition. They developed three models, in particular multilayer perceptron (MLP), LSTM, and biLSTM, with the latter achieving a significant higher accuracy. Also, the authors in [146] studied PPG signals eliciting traditional hand-crafted features from the time and frequency domain combined with a deep belief network (DBN), attaining over 65% accuracy in the 4-class pain estimation task. Utilizing kinematics data, Hu et al. [147] studied healthy and with low back pain (LBP) populations. Their approach was based on 2-stacked LSTM layers and achieved more than 97% accuracy in binary classification when fed with raw motion data. Finally, Mamontov et al. [148] reported the first study that adopted evolutionary algorithms and designed an optimized architecture of a recurrent neural network (RNN) and were used for pain estimation. Using this approach and EDA signals, they achieved 91.94% accuracy in a pain detection setting.

### 3.1.6. Audio-Based

A few studies have focused on using audio information to identify pain and/or estimate its intensity. These are shown in Table 11. Such approaches are particularly relevant for neonates since, due

**Table 7**
Studies utilizing vision-based methods and the temporal dimension of pain.

| Paper | Input | | | | Processing | | | | Evaluation | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Modality | Non deep features | Fusion M/E | Temporal Exploitation | Deep model | Non deep model | Learning Method | Classific. /Regres. | Objective | GT | Number subjects | Validation Method | Dataset | Metrics |
| '17 [86] | F (RGB) | HOG, distance metrics | - DF | NL | 2D CNN$^+$ | RVR | SL | R | IC | PS | 25 | LOSO | UNBC | 0.99 RMSE, 0.67 PCC |
| '17 [87] | F (RGB) | HOG, distance metrics | - DF | NL | 2D CNN$^+$ | RVM | SL | R | IC | PS | 25 | LOSO | UNBC | 1.04 RMSE, 0.64 PCC |
| '18 [88] | F (RGB) | - | - - | NL | 2D CNN | - | SL | R | IC | PS | 25 | LOSO | UNBC | 1.20 RMSE, 0.47 PCC |
| '18 [94] | F (RGB) | - | - - | I | 3D CNN | - | SL | R | IC | PS | 25 | LOSO | UNBC | 0.53 MSE, 0.84 PCC$^‡$ |
| '18 [95] | F (RGB) | HOG, geometric difference | - DF | I | 3D CNN | SVR | SL | R | IC | PS | 25 | LOSO | UNBC | 0.94 RMSE, 0.67 PCC |
| '20 [101] | F (RGB) | - | - - | I | 3D CNN$^+$ | - | WSL | R | IC | PS | 24 ? | LOSO | UNBC & RECOLA | 0.64 MAE, 0.82 PCC$^‡$ |
| '16 [118] | F (RGB) | - | - FF | E | RCNN | - | SL | R | IC | PS | 25 | LOSO | UNBC | 1.54 MSE, 0.65 PCC |
| '17 [119] | F (RGB) | - | - FF | E | [2D CNN$^+$, LSTM]$^\cup$ | - | SL | C, R | P, IC$^1$ | PS | 25 | LOSO | UNBC | 0.74$^1$ MSE, 0.78$^1$ PCC$^‡$ |
| '17 [120] | F (RGB) | - | - FF | E | [2D CNN$^+$, LSTM]$^\cup$ | - | SL | C | ID | PS | 25 | LOSO | UNBC | 61.90 ACC |
| '19 [121] | F (RGB) | - | - FF | E | [2D CNN$^+$, LSTM]$^\cup$ | - | SL | C | ID | PS | 25 | LOSO | UNBC | 75.20 ACC |
| '20 [122] | F (RGB) | PCA | - DF | E | [2D CNN$^+$, 1D CNN, biLSTM]$^\cup$ | - | SL | C | ID | PS | 25 | LOSO$^†$ | UNBC | 85.00 ACC$^‡$ |
| '19 [123] | F (RGB) | - | - - | E | [2D CNN$^+$, GRU]$^\cup$ | - | SL | C | ID, IC | PS | 25 | LOSO | UNBC | 85.40 ACC, 0.62 MSE$^‡$ |
| '19 [124] | F (RGB) | - | - FF | E | [2D CNN, RCNN]$^\cup$ | - | SL | R | IC | PS | 25 | LOSO | UNBC | 1.29 MSE, 0.73 PCC |
| '17 [129] | F (RGB) | - | - FF | E | biLSTM | HCRF, FC | SL | C | IC | O, S | 25 | hold-out | UNBC | 2.46 MAE$^‡$ |

**Non deep features:** PCA: principal component analysis **Temporal Exploitation:** NL: non-machine learning method I: implicit method E: explicit method **Deep models:** RCNN: recurrent convolutional neural network LSTM: long short memory networks biLSTM: bidirectional neural network GRU: gated recurrent unit **Non deep models:** SVM: support vector machine RVM: relevance vector machine GPM: Gaussian process regression model HCRF: hidden conditional random fields FC: fully connected SVR: support vector regression **Objective:** I2: intensity in binary pairs **Metrics:** PCC: Pearson correlation coefficient

**Table 8**
Studies utilizing vision-based methods and the temporal dimension of pain.

| Paper | Input | | | | Processing | | | | Evaluation | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Modality | Non deep features | Fusion M/E | Temporal Exploitation | Deep model | Non deep model | Learning Method | Classific. /Regres. | Objective | GT | Number subjects | Validation Method | Dataset | Metrics |
| '19 [33] | F (RGB) | HOG, distance metrics | - DF | NL | 2D CNN | RVR | SL | R | IC | O | 13 | LOSO | APN-DB | 1.71 MAE$^\ddagger$ |
| '19 [89] | F (RGB) | - | - - | NL | R-GAN | - | UL | C | genuine vs posed | PS, ST | 25, 34, 87, 87 | ? | UNBC & STOIC & BioVid (A) & BioVid (D) | 90.97 ACC |
| '20 [90] | F (RGB) | - | - FF | NL | 2D CNN$^+$ | - | SSL | C | IC | P, ST | 25, 87 | LOSO | UNBC$^1$, BioVid (A)$^2$ $\oplus$ | 0.78$^1$ PCC$^\ddagger$, 71.02$^2$AUC$^\ddagger$ |
| '21 [91] | F (RGB) | AUs intensity | - H | NL | 2D CNN$^+$ | RF | SL | C | ID | ST | 127 | k-fold | X-ITE | 25.00 ACC |
| '19 [93] | F (RGB) | - | - - | NL | 2D CNN | - | SL | C | P | ST | 87, 134 | k-fold | BioVid (A) & X-ITE$^\oplus$ | 67.90 ACC |
| '20 [130] | F (RGB) | - | - FF | E | [2D CNN$^+$, GRU]$^\cup$ | - | SL | R | IC | O, S | 25 | k-fold | UNBC | 2.34 MAE |
| '20 [132] | F (RGB) | - | - FF | E | [2D CNN$^+$, GRU]$^\cup$ | - | SL | R | IC | PS | 19 | LOSO | UNBC | 0.21 MSE, 0.89 PCC |
| '19 [133] | F (RGB) | - | - FF | E | [2D CNN, LSTM]$^\cup$ | - | SL | R | IC | PS | 24 | LOSO | UNBC | 1.22 MSE$^\ddagger$, 0.40 PCC$^\ddagger$ |
| '20 [136] | F (RGB) | AUs intensity | - - | E | LSTM | - | SL | R | IC | O | 36 | hold-out | EmoPain | 2.12 RMSE, 1.60 MAE$^\ddagger$ |
| '20 [138] | F (RGB) | - | - FF | E | [2D CNN$^+$, LSTM]$^\cup$ | - | SL | C | P | O | ? | k-fold | UNBC | 78.20 ACC$^\ddagger$ |
| '20 [142] | F (RGB) | - | - DF | E | [2D CNN, biLSTM, NN]$^\cup$ | - | SL | C | P | ST | 87, 40 | LOSO | BioVid (A)$^1$, SenseEmotion$^2$ | 69.25$^1$ ACC, 64.35$^2$ ACC |
| '20 [155] | F (RGB) | - | - FF | E | [2D CNN$^+$, GRU]$^\cup$ | - | SL | C | ID, IC | PS | 25 | LOSO | UNBC | 0.84 ACC, 0.69 PCC$^\ddagger$ |

$^\oplus$: The authors provide experiments with cross-dataset settings **Fusion:** H: hybrid **Non deep models:** RF: random forest classifier

**Table 9**
Studies utilizing vision-based methods and the temporal dimension of pain.

| Paper | Input | | | | Processing | | | | Evaluation | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Modality | Non deep features | Fusion M/E | Temporal Exploitation | Deep model | Non deep model | Learning Method | Classific. /Regres. | Objective | GT | Number subjects | Validation Method | Dataset | Metrics |
| '21 [97] | F (RGB) | facial landmarks | - DF | I | [3D CNN$^+$, 2D CNN$^+$, 1D CNN, FC]$^\cup$ | - | SL | R | IC | PS | 25 | LOSO | UNBC | 0.76 MSE, 0.82 PCC$^\ddagger$ |
| '19 [99] | F (RGB) | - | - - | I | [2D CNN$^+$, 3D CNN]$^\cup$ | - | UL, SL | C, R | IC$^1$, P$^2$ | P, ST | 25, 87 | LOSO | UNBC$^1$, BioVid (A)$^2$ | 0.92$^{11}$ PCC$^\ddagger$, 86.02$^{22}$ AUC |
| '20 [100] | F (RGB) | - | - - | I | 3D CNN$^+$ | - | WSL | R | IC | PS | 24,?, 87, 18 | LOSO | UNBC$^1$ & RECOLA & BioVid (A)$^2$ $\oplus$ | 0.74$^1$ PCC, 0.34$^2$ PCC |
| '20 [103] | F (RGB) | PCA | - FF | I | [2D CNN$^+$, TCN]$^\cup$ | - | SL | C | ID | P, ST | 25, 20 | LOSO$^\dagger$ | UNBC$^1$, MIntPAIN$^2$ | 92.44$^1$ ACC$^\ddagger$, 89.00$^2$ ACC$^\ddagger$ |
| '20 [104] | F (RGB) | - | - - | I | 2D CNN | - | SL | C, R | IC, P$^1$ | P | 95, 25 | k-fold | UofR & UNBC$^1$ | 82.00$^{11}$ PCC$^\ddagger$ |
| '20 [106] | F (RGB) | AUs occurrence | - FF | I | 1D CNN | - | SL | R | IC | P | 24, 87 | hold-out | UNBC$^1$, BioVid (A) | 0.80$^1$ CCC |
| '20 [125] | F (RGB) | PCA | - DF | E | [2D CNN$^+$, 1D CNN, biLSTM]$^\cup$ | - | SL | C | ID | PS, ST | 25, 20 | k-fold | UNBC$^1$, MIntPAIN$^2$ | 86.00$^1$ ACC$^\ddagger$ 92.26$^2$ ACC$^\ddagger$ |
| '20 [126] | F (RGB) | - | - FF | E | [2D CNN$^+$, LSTM]$^\cup$ | - | SL | R | P, IC$^1$ | O | 45 | LOSO | NPAD | 3.99$^1$ MSE, 1.55$^2$ MAE |
| '19 [127] | F (RGB) | - | - FF | E | [2D CNN$^+$, LSTM]$^\cup$ | - | UL | C | P | ST | 40 | LOSO | SenseEmotion | 60.61 ACC |
| '21 [131] | F (RGB) | - | - - | E | [2D CNN$^+$, LSTM]$^\cup$ | - | SL | R | IC | P | 25, 27 | LOSO | UNBC$^1$, DISFA$^\oplus$ | 0.60$^+$ MSE, 0.82$^+$ PCC$^\ddagger$ |
| '21 [134] | F (RGB) | - | - - | E | [2D CNN$^+$, Transformer]$^\cup$ | - | SL | R | IC | P | 25 | LOSO | UNBC | 0.40 MSE, 0.76 PCC$^\ddagger$ |
| '21 [137] | F (RGB) | - | - - | E | 2D C-LSTM | - | SL | C | ID | S | 29 | hold-out | other | 69.58 F1 |
| '19 [140] | F (RGB) | - | - FF | E | SLSTM | - | SL | C | P$^1$, ID$^2$ | ST | 85 | LOSO | BioVid (A) | 61.70$^1$ ACC 29.70$^2$ ACC |
| '21 [156] | F (RGB) | - | - - | I | 3D CNN$^+$ | - | SL | R | IC | S | 25 | k-fold | UNBC | 0.66 ICC$^\ddagger$ |

**Fusion:** H: hybrid **Deep models:** TCN: temporal convolutional neural network C-LSTM: convolutional-LSTM SLTM: sparse long short memory network **Learning Method:** SSL: self-supervised learning **Metrics:** F1: F1 score CCC: concordance correlation coefficient

**Table 10**
Studies utilizing contact-sensor' information.

| Paper | Input | | | | Processing | | | | Evaluation | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Modality | Non deep features | Fusion M/E | Temporal Exploitation | Deep model | Non deep model | Learning Method | Classific. /Regres. | Objective | GT | Number subjects | Validation Method | Dataset | Metrics |
| '20 [143] | EEG | - | - FF | I | 1D TCN | - | S | C | ID | S | 32 | k-fold | other | 97.30 ACC$^{\ddagger}$ |
| '20 [144] | EEG | - | - - | I | AE (TCN) | LR | UL, S | C | P | S | 29 | LOSO | other | 74.60 ACC |
| '21 [145] | fNIRS | - | - - | E | biLSTM | - | SL | C | ID | S | 18 | k-fold | other | 90.60 ACC$^{\ddagger}$ |
| '19 [146] | PPG | - | - - | NL | DBN | SBM | U, SL | C | P$^1$, ID$^2$ | S | 100 | k-fold | other | 86.79$^1$ ACC, 65.57$^2$ ACC |
| '18 [147] | kinematatics | - | - FF | E | LSTM | - | SL | C | P | LBP | 44 | LOSO | other | 97.20 ACC$^{\ddagger}$ |
| '19 [148] | EDA | - | - FF | E | [RNN, LSTM, GRU, NN]$^{\cup}$ | SelfCGA, selfCGP, PSOPB | SL | C | P | ST | 40 | LOSO | Sense- Emotion | 81.94 ACC |
| '21 [168] | EDA | - | - - | I | NN | - | SL | C | P$^1$, I2 | ST | 87, 55 | LOSO | BioVid (A)$^1$, PainMonit$^2$ | 84.22$^{11}$ ACC$^{\ddagger}$, 86.50$^{12}$ ACC$^{\ddagger}$ |

**Modality:** PPG: photoplethysmogram fNIRS: functional near-infrared spectroscopy EEG: electroencephalography EDA: electrodermal activity **Deep models:** DBN: Deep belief network RNN: recurrent neural network **Non deep models:** SBM: selective bagging model LR: Logistic Regression SelfCGA: Self-Configuring Genetic Algorithm SelfCGP: Self-Configuring Genetic Programming PSOPB: Particle Swarm Optimisation with parasitic behaviour **GT:** LBP: low back pain vs healthy population

**Table 11**
Studies utilizing audio information.

| Paper | Input | | | | Processing | | | | Evaluation | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Modality | Non deep features | Fusion M/E | Temporal Exploitation | Deep model | Non deep model | Learning Method | Classific. /Regres. | Objective | GT | Number subjects | Validation Method | Dataset | Metrics |
| '16 [149] | audio (cry) | - | - - | - | 2D CNN | - | SL | C | P | O | ? | k-fold | other | 78.50 ACC |
| '19 [150] | audio (cry) | - | - - | - | 2D CNN | - | SL | C | P | O | 31 | LOSO$^{\dagger}$ | NPAD | 96.77 ACC$^{\ddagger}$ |
| '19 [151] | audio (breathing) | MFCCs, RASTA-PLP, DTD | - FF | E | [2D CNN, LSTM]$^{\cup}$ | RFc | SL | C | P | ST | 40 | LOSO | Sense- Emotion | 64.39 ACC |
| '17 [152] | audio (voice) | prosodic-spectral features, SF | - FF | E | LSTM$^{+}$ | SVM | UL, SL | C | P$^1$, ID$^2$ | S | 63 | LOSO | other | 72.30$^1$ UAR, 54.20$^2$ UAR |

**Non deep features:** MFCCs: Mel Frequency Cepstral Coefficients RASTA-PLT: Relative Spectral Perceptual Linear Predictive DTD: descriptors from temporal domain SF: statistical features to the frequent facial and body occlusions that are observed, studying the cries is believed to be more appropriate way for detecting pain. Specifically, Chang and Li [149] focused on studying infants' cries in an attempt to distinguish the states of being hungry, in pain, and feeling sleepy. The authors converted the audio signals to 2D spectrograms through fast Fourier transform (FFT) and used these to train a 2D CNN as a feature extractor. Similarly, the authors of [150] utilized spectrograms derived from recorded sounds and employed an identical model as [81]. Also, Thiam and Schwenker [151] detected the presence of pain in adults through the analysis of breathing sounds. In so doing they exploited deep learned features resulting from spectrograms with Mel-scaled short-time Fourier transform and a plethora of handcrafted cues. A CNN followed by a biLSTM captured the spatial and temporal dependencies and concurrently combined low and high-level features. Tsai et al. [152] scrutinized pain events during emergency triage, developing a framework with an LSTM autoencoder to extract temporal features from verbal behavior, reporting promising results.

### 3.2. Multimodal approaches

Since pain is a multidimensional phenomenon, a promising direction is to combine modalities in a multimodal system. Heterogeneous information sources may complement each other and lead to improved specificity and sensitivity. Generally, as reported in [18], if the predictive performances of the single modalities are sufficiently good, their fusion tends to improve the results. Also, the utilization of cues that originate from diverse channels might prove not only helpful but also necessary, especially in clinical settings, where for various reasons, a modality may be unavailable (*e.g.,* the patient rotates, and his/her face is occluded). The information channels might rely on the following: (1) on the same hardware sensor but various regions of interest, *e.g.,* RGB facial images & RGB body images [153], (2) from different hardware sensors but the same region of interest, *e.g.,* RGB facial images & thermal facial images [29] or (3) different hardware sensor and information sources, *e.g.,* RGB facial images & ECG signals [154]. Table 12 presents the available studies employing multimodal approaches.

#### 3.2.1. Non-Temporal exploitation

One of the well-known and most exploited combinations of biosignals is the use of EDA, EMG, and ECG due to the fact that the key reference pain databases all include these information channels. Using these modalities, Thiam et al. [157] adopted an early fusion technique via concatenation, created a 2D representation, and fed it to a 9-layer 2D CNN. The results of their experiments show that EDA is highly correlated to pain intensity and that the fusion of the three information channels did not achieve a better performance compared to using EDA alone. Other authors employed least generative adversarial networks (LSGANs) to augment the EMG, EDA, and ECG samples [158], reporting that the classification performance of an SVM significantly improved, utilizing the augmented dataset. Haque et al. [29] introduced a new pain dataset *MIntPAIN* which, in addition to RGB videos, also includes depth and thermal videos, to be used for the multi-class (*i.e.,* 5 levels) recognition of pain. The authors merged the three visual modalities creating a 5D matrix (RGB+D+T), with which they fed the well-known pre-trained model *VGG-Face* [75]. Their experiments show that combining these three modalities produced a better classification performance.

#### 3.2.2. Temporal exploitation

Zhi et al. [159] presented a multimodal, stream-integrated neural network utilizing videos and biosignals. The authors combined raw facial video frames and optical flow images to capture spatio-temporal dependencies through 3D CNNs, which were integrated with the biosignals' features obtained from LSTMs. The whole network is trained in an end-to-end manner achieving good results, which were improved compared to their unimodal approaches. Besides the facial area, Salekin et al. [153] assessed neonatal pain from videos utilizing body movements. After detecting the specific areas, the video frames were inserted separately into a pre-trained *VGG-16* [42], which was connected to an LSTM to capture the dynamics from the frame sequences. The authors in [160] employed three information channels, specifically facial expressions, body movements, and crying sounds of neonatal. The results showed that the decision fusion of the modalities outperformed the unimodal approaches. The authors of [161] in addition to exploiting the EMG, EDA and ECG biosignals, they also experimented with combinations of handcrafted and learned features extracted from a biLSTM model. Initially, the minimum relevance method (MRMR) was applied to reduce the number of extracted features, obtaining good results. Some authors used deep denoising convolutional autoencoders (DDCAEs) for binary pain classification. In the work reported in [162], the latent representation was identified for each biopotential, followed by a weighing stage before the classification process. In a subsequent work [163], the same authors extended the DDCAE, by adding an attention mechanism that improved the performance. Furthermore, they experimented with self-supervised learning, which can reduce the training samples drastically, achieving very similar results. Subramaniam and Dass [164] explored the properties of EDA and ECG to investigate the differences in the pain manifestation between men and women. They designed a framework that included 4-convolutional layers followed by a 1-layer LSTM as a feature extractor. The usage of both biosignals gave the best results for pain discrimination. In parallel, it was shown that a significant performance difference exists between men and women, indicating that gender-based pain recognition is challenging yet possible.

Besides using EDA, EMG, and ECG, several other combinations of biosignals have been reported in the literature. Zhao et al. [165] employed information streams from PPG, EDA, and temperature signals. After the concatenation of the signals, they developed 2D convolutions to extract spatial features and employed time windows to exploit temporal information. Another study [166] reports successful pain estimation through whole-body MoCap sensors and EMG. They used an autoencoder integrating LSTM layers with an additional attention mechanism. The reported findings indicate that it is feasible to achieve acceptable performance while significantly reducing the training process time due to the raw data's latent space representation. Likewise, Li et al. [167] also used MoCap and EMG as information channels and experimented with several LSTMs configurations to predict the pain intensity. They report the best performance when a 3-layer vanilla LSTM combined with a 3-layer fully connected network was used.

## 4. Discussion

This section includes an analysis of the reviewed studies, focuses on answering the research questions established, provides a deeper analysis of the current approaches and their limitations, and provides suggestions for future research directions.

### 4.1. Input

Regarding the type of modalities used for the pain assessment, we observe a significant imbalance between unimodal and multimodal approaches. Specifically, more than 86% of reported studies are based on unimodal approaches, although the databases used in

**Table 12**
Studies utilizing multimodal approaches.

| Paper | Input | | | | Processing | | | | Evaluation | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Modality | Non deep features | Fusion M/E | Temporal Exploitation | Deep model | Non deep model | Learning Method | Classific. /Regres. | Objective | GT | Number subjects | Validation Method | Dataset | Metrics |
| ’18 [29] | F (RGB, thermal, depth) | - | RF - | - | 2D CNN$^{+}$ | - | SL | C | ID | S | 20 | k-fold | MIntPAIN | 36.55 ACC |
| ’19 [153] | F, B (RGB) | - | FF - | E | [2D CNN$^{+}$, LSTM]$^{\cup}$ | - | SL | C | P | O | 31 | LOSO | other | 92.48 ACC$^{\ddagger}$ |
| ’19 [154] | F (RGB), ECG, EDA | biosignals’ features$^{\oslash}$ | FF FF | - | 2D CNN$^{+}$ | RFc | SL | C | I2 | S | 85 | k-fold | BioVid (A) | 74.00 ACC |
| ’19 [157] | EDA, EMG, ECG | - | RF - | - | 2D CNN | - | SL | C | P$^{1}$ I2, ID$^{2}$ | S | 87, 86 | LOSO | BioVid (A)$^{1}$<br>BioVid (B) | 84.40$^{11}$ ACC$^{\ddagger}$, 36.54$^{12}$ ACC$^{\ddagger}$ |
| ’20 [158] | EDA, EMG, ECG | Boruta features | FF - | - | LSGAN | SVM | UL, SL | C | I2, ID$^{1}$ | S | 85 | hold-out | BioVid (A) | 82.80$^{1}$ ACC |
| ’21 [159] | F (RGB), EDA, EMG, ECG | optical flow | FF FF | NL, E, I | [3D CNN, LSTM]$^{\cup}$ | - | SL | C, R | P$^{1}$, I2, ID$^{2}$ | S | 87, 40 | k-fold$^{\dagger}$ | BioVid (A)$^{1}$, MIntPain | 68.20$^{11}$ ACC$^{\ddagger}$, 28.10$^{21}$ ACC |
| ’21 [160] | F, B (RGB), sound | - | DF - | E | [2D CNN$^{+}$, LSTM]$^{\cup}$ | - | SL | C | P | O | 45 | LOSO | NPAD | 78.95 ACC$^{\ddagger}$ |
| ’20 [161] | EDA, EMG, ECG | MRMR, biosignals’ features | RF FF | E | biLSTM | NN | SL | C | P$^{1}$, I2 | S | 87 | LOSO | BioVid (A) | 83.30$^{1}$ ACC |
| ’20 [162] | EDA, EMG, ECG | - | FF - | I | [DDCAE, NN]$^{\cup}$ | - | UL, SL | C | P$^{1}$, I2 | S | 87 | LOSO | BioVid (A) | 83.99$^{1}$ ACC$^{\ddagger}$ |
| ’21 [163] | EDA, EMG, ECG, RSP | - | FF - | I | [DDCAE, NN]$^{\cup}$ | - | UL, SL, SSL | C, R | P$^{1}$, ID$^{2}$, IC | S | 87, 40 | LOSO | BioVid (A)$^{1}$, Sense- Emotion | 84.25$^{11}$ ACC$^{\ddagger}$, 35.44$^{21}$ ACC$^{\ddagger}$ |
| ’21 [164] | EDA, ECG | - | FF - | E | 1D CNN, LSTM | - | UL | C | P$^{1}$, I2 | S | 67 | hold-out | BioVid (A) | 81.71$^{1}$ ACC |
| ’20 [165] | PPG, EDA, temperature | - | RF - | I | 2D CNN | - | SL | R$^{\circ}$ | P$^{1}$, ID$^{2}$ | S | 21 | k-fold | other | 96.30$^{1}$ ACC, 95.23$^{2}$ ACC |
| ’20 [166] | MoCap, EMG | - | RF - | E | AE, LSTM | - | UL, SL | C | ID | O | 23 | LOSO$^{\dagger}$ | EmoPain | 52.60 ACC$^{\ddagger}$ |
| ’20 [167] | MoCap, EMG | - | RF - | E | LSTM, NN | - | UL | C | ID | O | 30 | hold-out | EmoPain | 80.00 ACC$^{\ddagger}$ |
| ’21 [169] | MoCap, EMG | - | RF - | E | LSTM, NN | - | SL | C | ID | O | 30 | LOSO$^{\dagger}$ | EmoPain | 54.60 ACC$^{\ddagger}$ |

$\oslash$: Not specifically described $^{\circ}$: Ordinal **Modality** F: face region B: body region EMG: electromyography **Non deep features:** MRMR: Minimum Redundancy Maximum Relevance method **Deep models:** LSGAN: Least Square Generative Adversarial Networks these efforts include more than one information channel. Also, regarding unimodal approaches, the contact sensor-based and audio-based approaches are a minority, with only seven and four studies available, respectively, compared to 84 studies employing a vision-based approach.

Regarding the multimodal approaches, because only a limited number of studies (*i.e.,* fifteen in total) belong to this category, there are no clear inferences regarding the effectiveness of particular combination of modalities. However, there are indications about the value of using EDA sensors' information over other available biopotentials. These data indicate that researchers have focused primarily on visual information processing due to the implementation complexity of multimodal frameworks or the limited user-friendliness of contact sensors for their daily use in non-laboratory settings. It is evident, in our view, that further study and exploitation of diverse combinations of modalities is necessary to properly assess the potential of such modalities for the task at hand, *i.e.,* pain assessment. Furthermore, respecting the utilization of non-deep features, twenty-eight studies exploited them in order to enhance the deep learned representations.

Finally, regarding approaches that exploit the temporal dimension of information, we identified three main strategies which capitalize on the temporal information of video frames or signal segments; non-machine learning-based, machine learning-based (implicit), and machine learning-based (explicit). The non-machine learning-based approaches, for example, are based on motion history images [142] or temporal distillation [90], which are traditional computer vision techniques, are constitute more straightforward but less sophisticated methods. On the contrary, machine learning-based approaches [100,140] comprise richer temporal information about the corresponding modality and provide the flexibility to fine-tune them to our specific needs, *e.g.,* emphasize on specific video frames. From the studies reviewed, 55% exploited features with a temporal dimension; the most common approach involved explicit methods, specifically the LSTM models. It is our view that, considering the fact that several studies report a superior performance when the temporal dimension of the information is exploited as compared to the non-temporal dimension approaches, an increased focus on such approaches is required.

### *4.2. Processing*

Concerning the machine learning approaches, various models and techniques have been employed in pain estimation. CNN models have been the most popular approach and have been applied in most studies. Specifically, more than 75% of all studies have utilized either 1D, 2D, or 3D filters, indicating that the operation of convolution is currently the fundamental element of deep learning. These models are followed in popularity by sequential models, *i.e.,* RNN, GRU, LSTM, and biLSTM. In addition, almost half of the studies have utilized a pre-trained model to achieve the desired performance, which may indicate that the available pain databases are not optimal for training a deep learning model from scratch. Regarding the non-deep learning models, twenty-six studies have adopted such an approach as an auxiliary decision component of the extracted deep-learned features, with SVMs and shallow neural networks being the most popular choice. It is obvious, in our view, that there is room for adopting newer types of deep machine learning architectures. Particularly the transformer-based models appear to be best suited regarding the explicit temporal exploitation of modalities since they are well known for their state-of-the-art performances in many other AI research fields [170].

The learning method employed is primarily supervised, although sixteen papers adopted or experimented with a different method, *i.e.,* unsupervised [40,54,89,99,127,144,146,152,158,162,166], self-supervised [90,163], semi-supervised [40], weakly supervised [100,101], and federated [110]. It is our view that self-supervised learning is the most appropriate method due to the limited pain data resources available and should therefore be utilized more often by the research community.

Finally, it is worth reporting that most studies, *i.e.,* approximately 70%, approach pain assessment as a classification problem, not a regression problem. It is our view again that the latter is closest to reality due to the continuous nature of pain sensation.

### *4.3. Evaluation*

The main objective of the studies reviewed was *(i)* the estimation of the pain intensity in a discrete scale (*i.e.,* multi-class classification), *(ii)* the estimation of the pain intensity in a continuous scale, and *(iii)* estimating the presence or absence of pain (*i.e.,* binary classification). Twenty-five studies focused on the issue of pain detection rather than the estimation of pain intensity, which is, in our opinion, less important clinically since it does not provide adequate information relevant to the management of pain. It is also evident that, from an engineering standpoint, detecting the presence or absence of pain is a less complicated and demanding task.

It is also of interest that a small number of studies investigated the issue of pain estimation from a different perspective. In specific, a study [89] attempted to detect genuine pain over acted pain. Also, the authors in [152] studied the pain events in emerging triage rather than laboratory settings, while in [154], they examined the potential to achieve pain detection on IoT devices in real time. Furthermore, in [63,64], the authors attempted to overcome the occluded faces problem. Finally, according to sociodemographic and psychological determinants, the authors in [164] conducted experiments related to gender, whereas in [104], they studied pain estimation in elders suffering from dementia. The limited number of studies exploring the pain phenomenon beyond ordinary settings or considering a different context is, in our view, an indication of the limitations of current approaches regarding their applicability in real-life environments and circumstances, *i.e.,* clinics, hospitals, *etc.*

Regarding ground truth, several annotation types exist, such as self-reported ratings, FACS, and other observer scales. As previously mentioned, the extracted temporal features are crucial for accurate pain intensity estimation. Therefore, the temporal granularity of the ground truth is also essential. Several studies indicate that the PSPI scores are not considered objective pain metrics. For example, it is mentioned in [171] that the PSPI may be zero, although the person feels pain or, in other cases, there are no visible facial pain movements in low-intensity pain events. In addition, pain expressions can exist that are not described in the FACS system, such as raising eyebrows or opening the mouth [172]. Furthermore, PSPI scores do not consider head and body movements related to pain, something valuable, especially in newborns [173]. We suggest avoiding using the PSPI scores as ground truth for the aforementioned reasons. Instead, we recommend the adoption of self-reports and observer scales on the video-segment level.

Nearly 54% of the studies adopted the leave-one-subject-out (LOSO) validation method. This fact proves that the specific approach is widely thought of as being more objective and that it better supports the generalizability of the models. However, on a practical level, adopting LOSO may prove to be non-optimal, considering the models' size and the time required for their training. When the researchers utilize other validation methods, *e.g.,* k-fold, hold-out, *etc.,* it is necessary to consider the circumstances where consecutive, highly correlated frames from the same subject are used for training and validation, leading to flawed estimation. Finally, when the researchers employ their own validation/testing sets, as expected, it proved impossible to compare the results among studies, especially between classification and regression approaches. For these reasons, we believe it would be beneficial if specific evaluation protocols were developed for each publicly available database.

### 4.4. Pain databases

The availability of appropriate public databases is perhaps the single most significant element in studying and solving the problem of automatic pain assessment. In evaluating available datasets, one should consider several aspects, such as the number of subjects, and characteristics of subjects, *i.e.,* age, sex, health status, and race. Furthermore, the truth given needs to be objective and offers real insights into the painful situation of the subject [76].

In Fig. 1, we present the number of papers in relation to the pain database used in every study. It is evident from this figure that the UNBC and BioVid were the most used public datasets. It should be noted that the subject's age in UNBC is not recorded, which is known to be a factor that influences pain manifestation [174,175]. Also, whereas in BioVid, the age is documented, the older subjects are only 65 years old. This is important, since pain and pain management is a growing concern among citizens aged 65 and older [176]. A similar situation exists regarding other pain databases (*i.e.,* X-ITE [36], EmoPain [34], and SenseEmotion [35]). It is also known that age causes skin changes, (*e.g.,* texture, rigidity, and elasticity) that affect emotional face recognition tasks [177]. Various race-related factors also lead to erroneous pain estimation due to differences in the expression of the subjects [178]. It should be noted that one study, *i.e.,* Nerella et al. [117] mentioned that their model achieved low performance when tested with African American patients. We also found one study [104] dealing with the issue of pain estimation in elders with dementia. In conclusion, developing objective, automated, and generalizable deep learning-based pain assessment systems will only be feasible if representative and balanced data are available for training and external validation.

### 4.5. Interpretation

In recent years, AI models have begun to demonstrate state-of-the-art performances in nearly every scientific field and to outperform humans at specific diagnostic tasks [179]. However, AI solutions in general, and deep neural networks in particular, lack transparency, leading to the term "black box AI", referring to the fact that these models learn complex functions that are inaccessible and often incomprehensible to humans [180]. This non-transparency is a primary reason for the criticism that deep learning methods are receiving [181]. Several techniques have been developed that provide insights into how the models work, *e.g.,* visualizations, gradients-backpropagation emphasizing specific units, *etc.* Interested readers should refer to the recent review [182] about the explanatory techniques of deep learning.

Table 13 presents the various approaches implemented to provide an interpretation of the model's decision. It is evident that only a small percentage of studies, 20 out of 110, have implemented interpretation methods to explain how the model operates and on what features and elements it focuses. We would like to indicate at this point that interpretable machine learning is a useful umbrella term that captures the *"extraction of relevant knowledge from a machine-learning model concerning relationships either contained in data or learned by the model"* [183]. In summary, it is worth mentioning that: *(i)* 18% of the studies reviewed present an approach to support the interpretability of the model's decision, *(ii)* all of the methods implemented relate to studies that use facial images as an input modality, and *(iii)* approximately 50% of the studies were implemented from three specific research groups. These observations indicate that the issue of interpretability/explainability in the context of deep learning methods is an area that demands additional focus. This is especially true when automatically classifying pain severity levels.

### 4.6. Current challenges & future research

This section addresses open challenges in automatic pain assessment and suggest future research efforts to advance the field. Regarding the available pain databases, it is clear that several limitations exist. Most of the subjects' important demographic characteristics (*e.g.,* sex, gender, age) are absent. The narrow diversity of the subjects' race is also evident. However, it is known that the facial structure and emotional expressions of Caucasians differ from the Asian and African populations [184]. In addition, it would be beneficial to include social interactions influencing pain manifestation. For instance, to record the subjects while accompanied by a partner of the same and different gender [185]. Another vital component of an automatic pain assessment system, especially for infants or people with communication disabilities, is estimating the specific location of the pain source. Future databases need to include pain stimulations applied to several body locations. Furthermore, the existing databases incorporating visual information provide videos with low to medium-resolution and frame rates, unable to capture facial micro-expressions. Regarding the modalities, a limited number of databases include audio which could be a valuable information channel. Additionally, based on audio is feasible to extract linguistic features and adopt multimodal approaches, integrating natural language processing (NLP) methods. Similar research efforts have already existed in affective computing literature [186]. Also, present and future datasets must provide specific validation protocols to establish an objective and accurate comparison platform between scientific studies in the community.

Regarding the research efforts for automatic pain assessment from the engineering standpoint, we believe that attention to several issues is needed. First, developing multimodal approaches is required to establish effective systems with adequate capabilities. Beyond the superior estimation performances reported over unimodal methods, it is also necessary for real-world scenarios where a modality channel may often disappear. In addition, the exploitation of the temporal dimension of each modality is necessary. We strongly encourage the adoption of machine learning models or other techniques capable of incorporating the dynamic nature of pain. Further, multi-level and low-intensity pain estimation requires additional efforts to achieve improved results. Another open research topic is the relation of pain with other affective states, such as negative emotions, which presumably coexist in painful events. Identifying these emotions could be valuable, enabling better pain assessment. The existence of occlusions or poor illumination conditions in vision-based systems also demands further consideration. We strongly recommend that researchers investigate these or analogous conditions, even if the available databases do not encompass similar scenarios. Similarly, the real-time application of an automatic assessment system is critical. For this reason, we suggest that future studies include throughput measurements for the developed models, *e.g.,* the number of images per second in the inference phase. An additional essential aspect of every AI system is its generalization capabilities. Since there are several available pain databases, evaluating the trained system/model across diverse data could be valuable. Finally, adopting AI systems in the clinical domain requires explainability capabilities for its decisions. The development/adoption of methods supporting the interpretability of model decisions would greatly enhance the clinical translation of tools supporting automatic pain assessment.

**Table 13**
Interpretation approaches.

| Paper | Year | Modality | Method |
|---|---|---|---|
| [45] | 2021 | F (RGB) | visualization (saliency maps) |
| [49] | 2018 | F (RGB) | visualization (heat maps) |
| [51] | 2021 | F (RGB) | visualization (saliency map) |
| [54] | 2016 | F (RGB) | visualization (learned filters) |
| [55] | 2021 | F (RGB) | visualization (learned filters) |
| [56] | 2019 | F (RGB) | visualization (heat maps), values of learned weights |
| [59] | 2018 | F (RGB) | visualization (saliency maps) |
| [62] | 2021 | F (RGB) | visualization (attention maps) |
| [63] | 2021 | F (RGB) | visualization (saliency map) |
| [64] | 2021 | F (RGB) | visualization (activation maps) |
| [74] | 2020 | F (RGB) | visualization (pixels contributions) |
| [87] | 2017 | F (RGB) | visualization (average saliency map) |
| [89] | 2019 | F (RGB) | visualization (generated intermediate representation) |
| [104] | 2020 | F (RGB) | visualization (saliency maps) |
| [106] | 2020 | F (RGB) | weights per AU (contribution of AUs) |
| [115] | 2019 | F (RGB) | visualization (feature maps) |
| [116] | 2021 | F (RGB) | visualization (integrated gradients) |
| [131] | 2021 | F (RGB) | visualization (heatmaps) |
| [132] | 2020 | F (RGB) | visualization (attention maps), values of learned weights |
| [133] | 2019 | F (RGB) | visualization (attention maps) |

### 4.7. Potential threats to validity & limitations

The present SLR includes articles retrieved from three databases: Scopus, IEEE Xplore, ACM Digital Library and PubMed. There exist many potential databases which could also use in order to find original studies. However, it is difficult to imagine that studies on automatic pain assessment were not found in the specific databases or from the reference lists in the retrieved articles. In addition, as depicted in Fig. 3, access to two studies could not be obtained, and they are not, as a result, included in the review.

As mentioned in Section 3, in the present review we did not examine preprocessing methods and techniques, *e.g.,* face detection, alignment, *etc*. However, such processes could influence the final results of the pain estimation. Likewise, we do not discuss data augmentation techniques and strategies, which are very important for deep learning approaches. However, we would like to point out that most of the studies included in this review do not provide information on the data augmentation techniques used, if any.

Finally, in most cases, this review presents the highest performance of each study and the deep learning-based methods that provide this performance. However, several articles reviewed contain information on experiments with more than one approach. The results of such experiments may be valuable in the context of specific research studies, but the lack of available space did not permit us to include all this information.

## 5. Conclusions

Research on developing automatic methods supporting pain assessment has, to date, yielded many interesting results, promising ideas, and successful approaches. This is particularly true when employing deep learning methods, which, in many cases have shown that it is feasible to achieve great results. Despite the significant progress observed, there are several limitations and challenges associated with deep learning approaches; the development of such models requires specific hardware, and significant amounts of data are required to extract rich and valuable features. Also, the computational cost under specific situations is a trade-off between performance and speed. In addition, further investigation in enhancing the interpretability of the developed models is an area demanding increased attention because model uptake in the clinical domain requires at least some degree of explanation to persuade clinical users to its adoption. Finally, with respect to the pain assessment community, the development of harmonized evaluation protocols enabling the comparison among studies is, in our view, evident. In addition, the creation of balanced, publicly available databases of adequate size for the training, validation, and testing of AI-based models is essential.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgment

This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 945175 (Project: CARDIOCARE). This paper reflects only the author's view, and the Commission is not responsible for any use that may be made of the information it contains.

## Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2023.107365.

## References

[1] H. Merskey, D.G. Albe-Fessard, J.J. Bonica, A. Carmen, R. Dubner, F.W.L. Kerr, C.A. Pagni, Editorial: the need of a taxonomy, Pain 6 (3) (1979) 247–252, doi:10.1016/0304-3959(79)90046-0.

[2] T. Jackson, S. Thomas, V. Stabile, M. Shotwell, X. Han, K. McQueen, A systematic review and meta-analysis of the global burden of chronic pain without clear etiology in low- and middle-Income countries: trends in heterogeneous data and a proposal for new assessment methods, Anesth. Analg. 123 (3) (2016) 739–748, doi:10.1213/ANE.0000000000001389.

[3] Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017, Lancet 392 (10159) (2018) 1789–1858, doi:10.1016/S0140-6736(18)32279-7.

[4] D. Turk, R. Melzack, The measurement of pain and the assessment of people experiencing pain, Handbook of Pain Assessment, 2001.

[5] L. De Ruddere, R. Tait, Facing Others in Pain: Why Context Matters, Springer International Publishing, Cham, 2018, pp. 241–269.

[6] P. Dinakar, A.M. Stillman, Pathogenesis of pain, Semin. Pediatr. Neurol. 23 (3) (2016) 201–208, doi:10.1016/J.SPEN.2016.10.003.

[7] V.J. Dzau, P.A. Pizzo, Relieving pain in America: insights from an institute of medicine committee, JAMA 312 (15) (2014) 1507–1508, doi:10.1001/JAMA.2014.12986.

[8] W.F. Stewart, J.A. Ricci, E. Chee, D. Morganstein, R. Lipton, Lost productive time and cost due to common pain conditions in the US workforce, JAMA 290 (18) (2003) 2443–2454, doi:10.1001/JAMA.290.18.2443.

[9] D.J. Gaskin, P. Richard, The economic costs of pain in the United States, J. Pain 13 (8) (2012) 715–724, doi:10.1016/J.JPAIN.2012.03.009.

[10] Z. Hammal, J.F. Cohn, Automatic, Objective, and Efficient Measurement of Pain Using Automated Face Analysis, Springer International Publishing, Cham, 2018, pp. 121–146.

[11] B.G.S. Dekel, A. Gori, A. Vasarri, M.C. Sorella, G. Di Nino, R.M. Melotti, Medical evidence influence on inpatients and nurses pain ratings agreement, Pain Res. Manage. 2016 (2016), doi:10.1155/2016/9267536.

[12] K.M. Hoffman, S. Trawalter, J.R. Axt, M.N. Oliver, Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites, Proc. Natl. Acad. Sci. 113 (16) (2016) 4296–4301, doi:10.1073/PNAS.1516047113.

[13] C. Catal, On the application of genetic algorithms for test case prioritization: a systematic literature review, in: EAST '12: Proceedings of the 2nd International Workshop on Evidential Assessment of Software Technologies, Association for Computing Machinery, New York, NY, USA, 2012, pp. 9–14, doi:10.1145/2372233.2372238.

[14] K.M. Prkachin, Assessing pain by facial expression: facial expression as nexus, Pain Res. Manage. 14 (1) (2009) 53–58, doi:10.1155/2009/542964.

[15] G. Zamzmi, R. Kasturi, D. Goldgof, R. Zhi, T. Ashmeade, Y. Sun, A review of automated pain assessment in infants: features, classification tasks, and databases, IEEE Rev. Biomed. Eng. 11 (2018) 77–96, doi:10.1109/RBME.2017.2777907.

[16] Z. Chen, R. Ansari, D. Wilkie, Automated pain detection from facial expressions using FACS: a review, (2018). arXiv:1811.07988

[17] T. Hassan, D. Seus, J. Wollenberg, K. Weitz, M. Kunz, S. Lautenbacher, J.-U. Garbas, U. Schmid, Automatic detection of pain from facial expressions: a survey, IEEE Trans. Pattern Anal. Mach. Intell. (2019), doi:10.1109/tpami.2019.2958341. 1–1

[18] P. Werner, D. Lopez-Martinez, S. Walter, A. Al-Hamadi, S. Gruss, R. Picard, Automatic recognition methods supporting pain assessment: a survey, IEEE Trans. Affect. Comput. (2019), doi:10.1109/TAFFC.2019.2946774.

[19] R. M. Al-Eidan, H. Al-Khalifa, A. Al-Salman, Deep-learning-based models for pain recognition: asystematic review, Appl. Sci. 10 (17) (2020) 5984, doi:10.3390/app10175984.

[20] E.L. Garland, Pain processing in the human nervous system: aselective review of nociceptive and biobehavioral pathways, Prim. Care Clin. Office Pract. 39 (3) (2012) 561–571, doi:10.1016/J.POP.2012.06.013.

[21] A.C.d.C. Williams, K.D. Craig, Updating the definition of pain, Pain 157 (11) (2016) 2420–2423, doi:10.1097/j.pain.0000000000000613.

[22] D.A. Delgado, B.S. Lambert, N. Boutris, P.C. McCulloch, A.B. Robbins, M.R. Moreno, J.D. Harris, Validation of digital visual analog scale pain scoring with a traditional paper-based visual analog scale in adults, J. Am. Acad. Orthop.Surg. Glob. Res. Rev. 2 (3) (2018) e088, doi:10.5435/JAAOSGlobal-D-17-00088.

[23] M. Haefeli, A. Elfering, Pain assessment, Eur. Spine J. 15 Suppl 1 (Suppl 1) (2006) S17–24, doi:10.1007/s00586-005-1044-x.

[24] K.M. Prkachin, P.E. Solomon, The structure, reliability and validity of pain expression: evidence from patients with shoulder pain, Pain 139 (2) (2008) 267–274, doi:10.1016/j.pain.2008.04.010.

[25] J. Lawrence, D. Alcock, P. McGrath, J. Kay, S.B. MacMurray, C. Dulberg, The development of a tool to assess neonatal pain, Neonatal Netw. 12 (6) (1993) 59–66.

[26] D.E. Weissman, D.J. Haddox, Opioid pseudoaddiction–an iatrogenic syndrom, Pain 36 (3) (1989) 363–366, doi:10.1016/0304-3959(89)90097-3.

[27] P. Lucey, J.F. Cohn, K.M. Prkachin, P.E. Solomon, I. Matthews, Painful data: the UNBC-McMaster shoulder pain expression archive database, in: 2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, FG 2011, 2011, pp. 57–64, doi:10.1109/FG.2011.5771462.

[28] S. Walter, S. Gruss, H. Ehleiter, J. Tan, H.C. Traue, S. Crawcour, P. Werner, A. Al-Hamadi, A.O. Andrade, G.M. Da Silva, The biovid heat pain database: data for the advancement and systematic validation of an automated pain recognition, in: 2013 IEEE International Conference on Cybernetics, 2013, pp. 128–131, doi:10.1109/CYBConf.2013.6617456.

[29] M.A. Haque, R.B. Bautista, F. Noroozi, K. Kulkarni, C.B. Laursen, R. Irani, M. Bellantonio, S. Escalera, G. Anbarjafari, K. Nasrollahi, O.K. Andersen, E.G. Spaich, T.B. Moeslund, Deep multimodal pain recognition: a database and comparison of spatio-temporal visual modalities, in: 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 250–257, doi:10.1109/FG.2018.00044.

[30] S. Brahnam, C.-F. Chuang, F.Y. Shih, M.R. Slack, SVM Classification of Neonatal Facial Images of Pain, in: I. Bloch, A. Petrosino, A.G.B. Tettamanzi (Eds.), Fuzzy Logic and Applications, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 121–128.

[31] S. Brahnam, L. Nanni, S. McMurtrey, A. Lumini, R. Brattin, M. Slack, T. Barrier, Neonatal pain detection in videos using the iCOPEvid dataset and an ensemble of descriptors extracted from gaussian of local descriptors, Appl. Comput. Inform. (2019), doi:10.1016/j.aci.2019.05.003.

[32] G. Zamzmi, P. Chih-Yun, D. Goldgof, R. Kasturi, T. Ashmeade, Y. Sun, A comprehensive and context-sensitive neonatal pain assessment using computer vision, IEEE Trans. Affect. Comput. (2019), doi:10.1109/TAFFC.2019.2926710.

[33] J. Egede, M. Valstar, M.T. Torres, D. Sharkey, Automatic neonatal pain estimation: an acute pain in neonates database, in: 2019 8th International Conference on Affective Computing and Intelligent Interaction, ACII 2019, 2019, pp. 475–481, doi:10.1109/ACII.2019.8925480.

[34] M.S.H. Aung, S. Kaltwang, B. Romera-Paredes, B. Martinez, A. Singh, M. Cella, M. Valstar, H. Meng, A. Kemp, M. Shafizadeh, A.C. Elkins, N. Kanakam, A. de Rothschild, N. Tyler, P.J. Watson, A.C. de C Williams, M. Pantic, N. Bianchi-Berthouze, The automatic detection of chronic pain-related expression: requirements, challenges and the multimodal emopain dataset, IEEE Trans. Affect. Comput. 7 (4) (2016) 435–451, doi:10.1109/TAFFC.2015.2462830.

[35] M. Velana, S. Gruss, G. Layher, P. Thiam, Y. Zhang, D. Schork, V. Kessler, S. Meudt, H. Neumann, J. Kim, F. Schwenker, E. André, H.C. Traue, S. Walter, The senseemotion database: a multimodal database for the development and systematic validation of an automatic pain and emotion-recognition system, in: F. Schwenker, S. Scherer (Eds.), Multimodal Pattern Recognition of Social Signals in Human-Computer-Interaction, Springer International Publishing, Cham, 2017, pp. 127–139, doi:10.1007/978-3-319-59259-6_11.

[36] S. Gruss, M. Geiger, P. Werner, O. Wilhelm, H.C. Traue, A. Al-Hamadi, S. Walter, Multi-modal signals for analyzing pain responses to thermal and electrical stimuli, J. Vis. Exp. (146) (2019), doi:10.3791/59057.

[37] B.A. Kitchenham, S. Charters, Guidelines for performing Systematic Literature Reviews in Software Engineering, Technical Report EBSE 2007-001, Keele University and Durham University Joint Report, 2007.

[38] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, S.E. Brennan, R. Chou, J. Glanville, J.M. Grimshaw, A. Hróbjartsson, M.M. Lalu, T. Li, E.W. Loder, E. Mayo-Wilson, S. McDonald, L.A. McGuinness, L.A. Stewart, J. Thomas, A.C. Tricco, V.A. Welch, P. Whiting, D. Moher, The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, PLoS Med. 18 (3) (2021) e1003583, doi:10.1371/journal.pmed.1003583.

[39] D. Kirk, C. Catal, B. Tekinerdogan, Precision nutrition: a systematic literature review, Comput. Biol. Med. 133 (2021) 104365, doi:10.1016/j.compbiomed.2021.104365.

[40] H. Pedersen, Learning appearance features for pain detection using the UNBC-McMaster shoulder pain expression archive database, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 9163, 2015, pp. 128–136, doi:10.1007/978-3-319-20904-3_12.

[41] J.O. Egede, S. Song, T.A. Olugbade, C. Wang, A.C.D.C. Williams, H. Meng, M. Aung, N.D. Lane, M. Valstar, N. Bianchi-Berthouze, EmoPain challenge 2020: multimodal pain evaluation from facial and bodily expressions, in: 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), 2020, pp. 849–856, doi:10.1109/FG47880.2020.00078.

[42] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, International Conference on Learning Representations, ICLR, 2015.

[43] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2016-Decem, IEEE Computer Society, 2016, pp. 770–778, doi:10.1109/CVPR.2016.90. 1512.03385.

[44] R. Yang, X. Hong, J. Peng, X. Feng, G. Zhao, Incorporating high-level and low-level cues for pain intensity estimation, in: 2018 Proceedings - International Conference on Pattern Recognition, Vol. 2018-Augus, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 3495–3500, doi:10.1109/ICPR.2018.8545244.

[45] A. Semwal, N.D. Londhe, Computer aided pain detection and intensity estimation using compact CNN based fusion network, Appl. Soft Comput. 112 (2021) 107780, doi:10.1016/j.asoc.2021.107780.

[46] S.A.S. Lakshminarayan, S. Hinduja, S. Canavan, Three-level training of multi-head architecture for pain detection, in: G.-F. F. Struc V. (Ed.), Proceedings - 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2020, Institute of Electrical and Electronics Engineers Inc., 2020, pp. 839–843, doi:10.1109/FG47880.2020.00071.

[47] V.T. Huynh, H.J. Yang, G.S. Lee, S.H. Kim, Multimodality pain and related behaviors recognition based on attention learning, in: G.-F. F. Struc V. (Ed.), Proceedings - 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2020, Institute of Electrical and Electronics Engineers Inc., 2020, pp. 814–818, doi:10.1109/FG47880.2020.00034.

[48] A. Semwal, N.D. Londhe, Automated Pain Severity Detection Using Convolutional Neural Network, in: R.V.S.N.M.N., S.K. Niranjan, V. Desai (Eds.), Proceedings of the International Conference on Computational Techniques, Electronics and Mechanical Systems, CTEMS 2018, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 66–70, doi:10.1109/CTEMS.2018.8769123.

[49] M. Tavakolian, A. Hadid, Deep binary representation of facial expressions: a novel framework for automatic pain intensity recognition, in: Proceedings - International Conference on Image Processing, ICIP, IEEE Computer Society, 2018, pp. 1952–1956, doi:10.1109/ICIP.2018.8451681.

[50] D. Yi, Z. Lei, S. Liao, S.Z. Li, Learning face representation from scratch, 2014, 1411.7923

[51] A. Semwal, N.D. Londhe, ECCNET: an ensemble of compact convolution neural network for pain severity assessment from face images, in: Proceedings of the Confluence 2021: 11th International Conference on Cloud Computing, Data Science and Engineering, 2021, pp. 761–766, doi:10.1109/Confluence51648.2021.9377197.

[52] V.J. Lawhern, A.J. Solon, N.R. Waytowich, S.M. Gordon, C.P. Hung, B.J. Lance, EEGNet: a compact convolutional neural network for EEG-based brain-

computer interfaces, J. Neural Eng. 15 (5) (2018), doi:10.1088/1741-2552/aace8c.
[53] C. Szegedy, S. Ioffe, V. Vanhoucke, A.A. Alemi, Inception-v4, inception-ResNet and the impact of residual connections on learning, in: 31st AAAI Conference on Artificial Intelligence, AAAI 2017, AAAI press, 2017, pp. 4278–4284. 1602.07261.
[54] R. Kharghanian, A. Peiravi, F. Moradi, Pain detection from facial images using unsupervised feature learning approach, in: 2016 Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, Vol. 2016-Octob, Institute of Electrical and Electronics Engineers Inc., 2016, pp. 419–422, doi:10.1109/EMBC.2016.7590729.
[55] R. Kharghanian, A. Peiravi, F. Moradi, A. Iosifidis, Pain detection using batch normalized discriminant restricted Boltzmann machine layers, J. Vis. Commun. Image Represent. 76 (2021), doi:10.1016/j.jvcir.2021.103062.
[56] D. Huang, Z. Xia, L. Li, K. Wang, X. Feng, Pain-awareness multistream convolutional neural network for pain estimation, J. Electron. Imaging 28 (04) (2019) 1, doi:10.1117/1.jei.28.4.043008.
[57] X. Xin, X. Lin, S. Yang, X. Zheng, Pain intensity estimation based on a spatial transformation and attention CNN, PLoS ONE 15 (8 August 2020) (2020) 1–15, doi:10.1371/journal.pone.0232412.
[58] S. Cui, D. Huang, Y. Ni, X. Feng, Multi-scale regional attention networks for pain estimation, in: 2021 13th International Conference on Bioinformatics and Biomedical Technology, Association for Computing Machinery, 2021, pp. 1–8, doi:10.1145/3473258.3473259.
[59] C. Li, Z. Zhu, Y. Zhao, Saliency supervision: an intuitive and effective approach for pain intensity regression, Vol. 11307, 2018, pp. 455–464, doi:10.1007/978-3-030-04239-4_41.
[60] F. Schroff, D. Kalenichenko, J. Philbin, FaceNet: a unified embedding for face recognition and clustering, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, 2015, pp. 815–823, doi:10.1109/CVPR.2015.7298682.
[61] X. Peng, D. Huang, H. Zhang, Pain intensity recognition via multi-scale deep network, IET Image Proc. 14 (8) (2020) 1645–1652, doi:10.1049/iet-ipr.2019.1448.
[62] X. Xin, X. Li, S. Yang, X. Lin, X. Zheng, Pain expression assessment based on a locality and identity aware network, IET Image Proc. 15 (12) (2021) 2948–2958, doi:10.1049/ipr2.12282.
[63] A. Semwal, N.D. Londhe, S-PANET: a shallow convolutional neural network for pain severity assessment in uncontrolled environment, in: 2021 IEEE 11th Annual Computing and Communication Workshop and Conference, CCWC 2021, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 800–806, doi:10.1109/CCWC51732.2021.9376052.
[64] A. Semwal, N.D. Londhe, MVFNet: a multi-view fusion network for pain intensity assessment in unconstrained environment, Biomed. Signal Process. Control 67 (2021), doi:10.1016/j.bspc.2021.102537.
[65] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 07-12-June, IEEE Computer Society, 2015, pp. 1–9, doi:10.1109/CVPR.2015.7298594.
[66] J.S. Lee, C.W. Wang, Facial pain intensity estimation for ICU patient with partial occlusion coming from treatment, in: 3rd International Conference on Biological Information and Biomedical Engineering, BIBE 2019, VDE Verlag GmbH, 2019, pp. 106–109. https://ieeexplore.ieee.org/abstract/document/8903345
[67] R.A. Virrey, W. Caesarendra, M.I. Bin Pg Hj Petra, E. Abas, A. Husaini, C. De Silva Liyanage, Milestone of pain intensity evaluation from facial action units, in: ICECOS 2019 - 3rd International Conference on Electrical Engineering and Computer Science, Proceeding, Institute of Electrical and Electronics Engineers Inc., 2019, pp. 54–57, doi:10.1109/ICECOS47637.2019.8984464.
[68] H. Nugroho, D. Harmanto, H.R. Hassan Al-Absi, On the development of smart home care: application of deep learning for pain detection, in: 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), 2019, pp. 612–616, doi:10.1109/iecbes.2018.8626710.
[69] F. Schroff, D. Kalenichenko, J. Philbin, FaceNet: a unified embedding for face recognition and clustering, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2015, pp. 815–823, doi:10.1109/CVPR.2015.7298682.
[70] L. Dai, J. Broekens, K.P. Truong, Real-time pain detection in facial expressions for health robotics, in: 2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos, ACIIW 2019, Institute of Electrical and Electronics Engineers Inc., 2019, pp. 277–283, doi:10.1109/ACIIW.2019.8925192.
[71] G. Menchetti, Z. Chen, D.J. Wilkie, R. Ansari, Y. Yardimci, A.E. Cetin, Pain detection from facial videos using two-stage deep learning, in: GlobalSIP 2019 - 7th IEEE Global Conference on Signal and Information Processing, Proceedings, Institute of Electrical and Electronics Engineers Inc., 2019, doi:10.1109/GlobalSIP45357.2019.8969274.
[72] C. Andrade, Internal, external, and ecological validity in research design, conduct, and evaluation, Indian J. Psychol. Med. 40 (5) (2018) 498–499, doi:10.4103/IJPSYM.IJPSYM_334_18. PMID: 30275631
[73] D. Liu, P. Fengjiao, O.O. Rudovic, R. Picard, DeepFaceLIFT: interpretable personalized models for automatic estimation of self-reported pain, in: N. Lawrence, M. Reid (Eds.), Proceedings of IJCAI 2017 Workshop on Artificial Intelligence in Affective Computing, Proceedings of Machine Learning Research, Vol. 66, PMLR, 2017, pp. 1–16. https://proceedings.mlr.press/v66/liu17a.html
[74] X. Xu, J.S. Huang, V.R. De Sa, Pain evaluation in video using extended multitask learning from multidimensional measurements, in: Proceedings of the Machine Learning for Health NeurIPS Workshop, Vol. 116, PMLR, 2020, pp. 141–154. https://proceedings.mlr.press/v116/xu20a.html
[75] O.M. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition, in: Proceedings of the British Machine Vision Conference (BMVC), British Machine Vision Association and Society for Pattern Recognition, 2015, pp. 41.1–41.12, doi:10.5244/c.29.41.
[76] P. Casti, A. Mencattini, M.C. Comes, G. Callari, D. Di Giuseppe, S. Natoli, M. Dauri, E. Daprati, E. Martinelli, Calibration of vision-based measurement of pain intensity with multiple expert observers, IEEE Trans. Instrum. Meas. 68 (7) (2019) 2442–2450, doi:10.1109/TIM.2019.2909603.
[77] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Advances in Neural Information Processing Systems 25 (NIPS 2012), 2012, doi:10.1145/3065386.
[78] L. Celona, L. Manoni, Neonatal facial pain assessment combining hand-crafted and deep features, Lect. Notes Comput. Sci. 10590 (2017) 197–204, doi:10.1007/978-3-319-70742-6_19.
[79] G. Levi, T. Hassner, Emotion recognition in the wild via convolutional neural networks and mapped binary patterns, in: ICMI 2015 - Proceedings of the 2015 ACM International Conference on Multimodal Interaction, ACM, New York, NY, USA, 2015, pp. 503–510, doi:10.1145/2823327.2823333.
[80] G. Lu, Q. Hao, K. Kong, J. Yan, H. Li, X. Li, Deep convolutional neural networks with transfer learning for neonatal pain expression recognition, in: X.G.N.X.L.K.L.M., Z. Xiao, L. Wang (Eds.), 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Institute of Electrical and Electronics Engineers Inc., 2018, pp. 251–256, doi:10.1109/FSKD.2018.8687129.
[81] G. Zamzmi, R. Paul, M.S. Salekin, D. Goldgof, R. Kasturi, T. Ho, Y. Sun, Convolutional neural networks for neonatal pain assessment, IEEE Trans. Biom. Behav.Identity Sci. 1 (3) (2019) 192–200, doi:10.1109/tbiom.2019.2918619.
[82] G. Zamzmi, D. Goldgof, R. Kasturi, Y. Sun, Neonatal pain expression recognition using transfer learning, (2018). arXiv preprint arXiv:1807.01631
[83] K. Chatfield, K. Simonyan, A. Vedaldi, A. Zisserman, Return of the devil in the details: delving deep into convolutional nets, 2014. 1405.353110.5244/c.28.6
[84] L. Celona, S. Brahnam, S. Bianco, Getting the most of few data for neonatal pain assessment, in: ACM International Conference Proceeding Series, Association for Computing Machinery, 2019, pp. 298–301, doi:10.1145/3329189.3329219.
[85] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein GAN, arXiv (2017). 1701.07875
[86] J. Egede, M. Valstar, B. Martinez, Fusing deep learned and hand-crafted features of appearance, shape, and dynamics for automatic pain estimation, in: 12th IEEE International Conference on Automatic Face & Gesture Recognition, 2017, pp. 689–696, doi:10.1109/FG.2017.87.
[87] J.O. Egede, M. Valstar, Cumulative attributes for pain intensity estimation, in: ICMI 2017 - Proceedings of the 19th ACM International Conference on Multimodal Interaction, Vol. 2017-Janua, Association for Computing Machinery, Inc., 2017, pp. 146–153, doi:10.1145/3136755.3136789.
[88] S. Jaiswal, J. Egede, M. Valstar, Deep learned cumulative attribute regression, in: Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 715–722, doi:10.1109/FG.2018.00113.
[89] M. Tavakolian, C.G.B. Cruces, A. Hadid, Learning to detect genuine versus posed pain from facial expressions using residual generative adversarial networks, in: Proceedings - 14th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2019, Institute of Electrical and Electronics Engineers Inc., 2019, doi:10.1109/FG.2019.8756540.
[90] M. Tavakolian, M. Bordallo Lopez, L. Liu, Self-supervised pain intensity estimation from facial videos via statistical spatiotemporal distillation, Pattern Recognit. Lett. 140 (2020) 26–33, doi:10.1016/j.patrec.2020.09.012.
[91] E. Othman, P. Werner, F. Saxen, A. Al-Hamadi, S. Gruss, S. Walter, Automatic vs. human recognition of pain intensity from facial expression on the X-ITE pain database, Sensors 21 (9) (2021), doi:10.3390/s21093273.
[92] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, MobileNetV2: inverted residuals and linear bottlenecks, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520. arXiv preprint arXiv:1801.04381.
[93] E. Othman, P. Werner, F. Saxen, A. Al-Hamadi, S. Walter, Cross-database evaluation of pain recognition from facial video, in: 2019 International Symposium on Image and Signal Processing and Analysis, ISPA, Vol. 2019-Septe, IEEE Computer Society, 2019, pp. 181–186, doi:10.1109/ISPA.2019.8868562.
[94] M. Tavakolian, A. Hadid, Deep spatiotemporal representation of the face for automatic pain intensity estimation, in: 2018 24th International Conference on Pattern Recognition (ICPR), Vol. 2018-Augus, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 350–354, doi:10.1109/ICPR.2018.8545324.
[95] J. Wang, H. Sun, Pain intensity estimation using deep spatiotemporal and handcrafted features, IEICE Trans. Inf. Syst. E101D (6) (2018) 1572–1580, doi:10.1587/transinf.2017EDP7318.
[96] D. Tran, L. Bourdev, R. Fergus, L. Torresani, M. Paluri, Learning spatiotemporal features with 3D convolutional networks, in: 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 4489–4497, doi:10.1109/ICCV.2015.510.
[97] Y. Huang, L. Qing, S. Xu, L. Wang, Y. Peng, HybNet: a hybrid network structure for pain intensity estimation, Vis. Comput. (2021), doi:10.1007/s00371-021-02056-y.
[98] S. Xie, C. Sun, J. Huang, Z. Tu, K. Murphy, Rethinking spatiotemporal feature learning: Speed-accuracy trade-offs in video classification, in: Com-

puter Vision – ECCV 2018, LNCS, Vol. 11219, 2018, pp. 318–335, doi:10.1007/978-3-030-01267-0_19.

[99] M. Tavakolian, A. Hadid, A spatiotemporal convolutional neural network for automatic pain intensity estimation from facial dynamics, Int. J. Comput. Vis. 127 (10) (2019) 1413–1425, doi:10.1007/s11263-019-01191-3.

[100] G.P. R, E. Granger, P. Cardinal, Deep domain adaptation for ordinal regression of pain intensity estimation using weakly-labeled videos, CoRR (2020) arXiv preprint arXiv:2010.15675.

[101] R. Gnana Praveen, E. Granger, P. Cardinal, Deep weakly supervised domain adaptation for pain localization in videos, in: G.-F. F. Struc V. (Ed.), 15th IEEE International Conference on Automatic Face and Gesture Recognition, Institute of Electrical and Electronics Engineers Inc., 2020, pp. 473–480, doi:10.1109/FG47880.2020.00139.

[102] J. Carreira, A. Zisserman, Quo vadis, action recognition? A new model and the kinetics dataset, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), https://arxiv.org/abs/1705.07750v3.

[103] G. Bargshady, X. Zhou, R.C. Deo, J. Soar, F. Whittaker, H. Wang, The modeling of human facial pain intensity based on temporal convolutional networks trained with video frames in HSV color space, Appl. Soft Comput. J. 97 (2020), doi:10.1016/j.asoc.2020.106805.

[104] S. Rezaei, A. Moturu, S. Zhao, K.M. Prkachin, T. Hadjistavropoulos, B. Taati, Unobtrusive pain monitoring in older adults with dementia using pairwise and contrastive training, IEEE J. Biomed. Health Inform. 25 (5) (2021) 1450–1462, doi:10.1109/JBHI.2020.3045743.

[105] H. GE, Training products of experts by minimizing contrastive divergence, Neural Comput. 14 (8) (2002) 1771–1800, doi:10.1162/089976602760128018.

[106] V. Pandit, M. Schmitt, N. Cummins, B. Schuller, I see it in your eyes: training the shallowest-possible CNN to recognise emotions and pain from muted web-assisted in-the-wild video-chats in real-time, Inform. Process. Manage. 57 (6) (2020) 102347, doi:10.1016/j.ipm.2020.102347.

[107] F. Wang, X. Xiang, C. Liu, T.D. Tran, A. Reiter, G.D. Hager, H. Quon, J. Cheng, A.L. Yuille, Regularizing face verification nets for pain intensity regression, in: 2017 Proceedings - International Conference on Image Processing, ICIP, Vol. 2017-Septe, IEEE Computer Society, 2018, pp. 1087–1091, doi:10.1109/ICIP.2017.8296449.

[108] M.C. Dragomir, C. Florea, V. Pupezescu, Automatic subject independent pain intensity estimation using a deep learning approach, in: 2020 8th E-Health and Bioengineering Conference, EHB 2020, 2020, pp. 1–4, doi:10.1109/EHB50910.2020.9280190.

[109] A. Semwal, N.D. Londhe, Automated facial expression based pain assessment using deep convolutional neural network, in: Proceedings of the 3rd International Conference on Intelligent Sustainable Systems, ICISS 2020, Institute of Electrical and Electronics Engineers Inc., 2020, pp. 366–370, doi:10.1109/ICISS49785.2020.9316099.

[110] O. Rudovic, N. Tobis, S. Kaltwang, B. Schuller, D. Rueckert, J.F. Cohn, R.W. Picard, Personalized federated deep learning for pain estimation from face images, (2021). arXiv:2101.04800

[111] K. Pikulkaew, E. Boonchieng, W. Boonchieng, V. Chouvatut, Pain detection using deep learning with evaluation system, in: Advances in Intelligent Systems and Computing, Vol. 1184, Springer Singapore, 2021, pp. 426–435, doi:10.1007/978-981-15-5859-7_42.

[112] S. El Morabit, A. Rivenq, M.-E.-N. Zighem, A. Hadid, A. Ouahabi, A. Taleb-Ahmed, Automatic pain estimation from facial expressions: acomparative analysis using off-the-shelf cnn architectures, Electronics 10 (16) (2021), doi:10.3390/electronics10161926.

[113] C. Li, A. Pourtaherian, L. Van Onzenoort, W. Ten, P.H.N. De With, Infant facial expression analysis: towards a real-time video monitoring system using R-CNN and HMM, IEEE J. Biomed. Health Inform. 25 (5) (2021) 1429–1440, doi:10.1109/JBHI.2020.3037031.

[114] N. Rathee, S. Pahal, P. Sheoran, Pain detection from facial expressions using domain adaptation technique, Pattern Anal. Appl. (2021), doi:10.1007/s10044-021-01025-4.

[115] G. Zamzmi, R. Paul, D. Goldgof, R. Kasturi, Y. Sun, Pain assessment from facial expression: neonatal convolutional neural network (N-CNN), in: 2019 Proceedings of the International Joint Conference on Neural Networks, Vol. 2019-July, Institute of Electrical and Electronics Engineers Inc., 2019, doi:10.1109/IJCNN.2019.8851879.

[116] L.P. Carlini, L.A. Ferreira, G.S. Coutrin, V.V. Varoto, T.M. Heiderich, R.X. Balda, M.M. Barros, R. Guinsburg, C.E. Thomaz, A convolutional neural network-based mobile application to bedside neonatal pain assessment, in: Conference on Graphics, Patterns and Images (SIBGRAPI), IEEE Computer Society, Los Alamitos, CA, USA, 2021, pp. 394–401, doi:10.1109/SIBGRAPI54419.2021.00060.

[117] S. Nerella, J. Cupka, M. Ruppert, P. Tighe, A. Bihorac, P. Rashidi, Pain action unit detection in critically ill patients, in: 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), 2021, pp. 645–651, doi:10.1109/COMPSAC51774.2021.00094.

[118] J. Zhou, X. Hong, F. Su, G. Zhao, Recurrent convolutional neural network regression for continuous pain intensity estimation in video, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2016, pp. 1535–1543, doi:10.1109/CVPRW.2016.191.

[119] P. Rodriguez, G. Cucurull, J. Gonalez, J.M. Gonfaus, K. Nasrollahi, T.B. Moeslund, F.X. Roca, Deep pain: exploiting long short-term memory networks for facial expression classification, IEEE Trans. Cybern. (2017), doi:10.1109/TCYB.2017.2662199.

[120] M. Bellantonio, M.A. Haque, P. Rodriguez, K. Nasrollahi, T. Telve, S. Escarela, J. Gonzalez, T.B. Moeslund, P. Rasti, G. Anbarjafari, Spatio-temporal pain recognition in CNN-based super-resolved facial images, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) LNCS, Vol. 10165, 2017, pp. 151–162, doi:10.1007/978-3-319-56687-0_13.

[121] G. Bargshady, J. Soar, X. Zhou, R.C. Deo, F. Whittaker, H. Wang, A joint deep neural network model for pain recognition from face, in: 2019 IEEE 4th International Conference on Computer and Communication Systems, ICCCS 2019, 2019, pp. 52–56, doi:10.1109/CCOMS.2019.8821780.

[122] G. Bargshady, X. Zhou, R.C. Deo, J. Soar, F. Whittaker, H. Wang, Enhanced deep learning algorithm development to detect pain intensity from facial expression images, Expert Syst. Appl. 149 (2020), doi:10.1016/j.eswa.2020.113305.

[123] A. Mauricio, F. Cappabianco, A. Veloso, G. Cámara, A sequential approach for pain recognition based on facial representations, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) LNCS, Vol. 11754, 2019, pp. 295–304, doi:10.1007/978-3-030-34995-0_27.

[124] S. Thuseethan, S. Rajasegarar, J. Yearwood, Deep hybrid spatiotemporal networks for continuous pain intensity estimation, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) LNCS, Vol. 11955, 2019, pp. 449–461, doi:10.1007/978-3-030-36718-3_38.

[125] G. Bargshady, X. Zhou, R.C. Deo, J. Soar, F. Whittaker, H. Wang, Ensemble neural network approach detecting pain intensity from facial expressions, Artif. Intell. Med. 109 (2020), doi:10.1016/j.artmed.2020.101954.

[126] M.S. Salekin, G. Zamzmi, D. Goldgof, R. Kasturi, T. Ho, Y. Sun, First investigation into the use of deep learning for continuous assessment of neonatal postoperative pain, in: G.-F. F. Struc V. (Ed.), Proceedings - 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2020, Institute of Electrical and Electronics Engineers Inc., 2020, pp. 415–419, doi:10.1109/FG47880.2020.00082.

[127] N. Kalischek, P. Thiam, P. Bellmann, F. Schwenker, Deep domain adaptation for facial expression analysis, in: 2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos, ACIIW 2019, Institute of Electrical and Electronics Engineers Inc., 2019, pp. 317–323, doi:10.1109/ACIIW.2019.8925055.

[128] G. French, M. Mackiewicz, M. Fisher, Self-ensembling for visual domain adaptation, in: 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings, 2018.

[129] D.L. Martinez, O. Rudovic, R. Picard, Personalized automatic estimation of self-reported pain intensity from facial expressions, in: 2017 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Vol. 2017-July, IEEE Computer Society, 2017, pp. 2318–2327, doi:10.1109/CVPRW.2017.286.

[130] D. Erekat, Z. Hammal, M. Siddiqui, H. Dibeklioglu, Enforcing multilabel consistency for automatic spatio-temporal assessment of shoulder pain intensity, in: ICMI 2020 Companion - Companion Publication of the 2020 International Conference on Multimodal Interaction, Association for Computing Machinery, Inc, 2020, pp. 156–164, doi:10.1145/3395035.3425190.

[131] M.T. Vu, M. Beurton-Aimar, P.-y. Dezaunay, M.C. Eslous, Automated pain estimation based on facial action units from multi-databases, in: Joint International Conference on Informatics, Electronics Vision (ICIEV) and International Conference on Imaging, Vision Pattern Recognition (icIVPR), 2021, pp. 1–8, doi:10.1109/ICIEVicIVPR52578.2021.9564244.

[132] D. Huang, Z. Xia, J. Mwesigye, X. Feng, Pain-attentive network: a deep spatio-temporal attention model for pain estimation, Multimed Tools. Appl. 79 (37–38) (2020) 28329–28354, doi:10.1007/s11042-020-09397-1.

[133] J. Yu, T. Kurihara, S. Zhan, Frame by frame pain estimation using locally spatial attention learning, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) LNCS, Vol. 11868, 2019, pp. 229–238, doi:10.1007/978-3-030-31321-0_20.

[134] H. Xu, M. Liu, A deep attention transformer network for pain estimation with facial expression video, in: J. Feng, J. Zhang, M. Liu, Y. Fang (Eds.), Biometric Recognition, Springer International Publishing, Cham, pp. 112–119. 10.1007/978-3-030-86608-2_13

[135] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: NIPS'17:" Proceedings of the 31st International Conference on Neural Information Processing Systems, Curran Associates Inc., Red Hook, NY, USA, 2017, pp. 5998–6008.

[136] A. Mallol-Ragolta, S. Liu, N. Cummins, B. Schuller, A curriculum learning approach for pain intensity recognition from facial expressions, in: G.-F. F. Struc V. (Ed.), 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), Institute of Electrical and Electronics Engineers Inc., 2020, pp. 829–833, doi:10.1109/FG47880.2020.00083.

[137] Y. Guo, L. Wang, Y. Xiao, Y. Lin, A personalized spatial-temporal cold pain intensity estimation model based on facial expression, IEEE J. Transl. Eng. Health Med. 9 (2021), doi:10.1109/JTEHM.2021.3116867.

[138] S. Rasipuram, B.N. Sai, D.B. Jayagopi, A. Maitra, Using deep 3D features and an LSTM based sequence model for automatic pain detection in the wild, in: G.-F. F. Struc V. (Ed.), 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), Institute of Electrical and Electronics Engineers Inc., 2020, pp. 781–785, doi:10.1109/FG47880.2020.00097.

[139] F.J. Chang, A. Tuan Tran, T. Hassner, I. Masi, R. Nevatia, G. Medioni, ExpNet: landmark-free, deep, 3D facial expressions, in: 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018, Institute of Elec-

trical and Electronics Engineers Inc., 2018, pp. 122–129, doi:10.1109/FG.2018.00027. 1802.00542.

[140] R. Zhi, M. Wan, Dynamic facial expression feature learning based on sparse RNN, in: Proceedings of 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference, ITAIC 2019, Institute of Electrical and Electronics Engineers Inc., 2019, pp. 1373–1377, doi:10.1109/ITAIC.2019.8785844.

[141] T. Blumensath, M. Davies, Iterative thresholding for sparse approximations, J. Fourier Anal. Appl. 14 (2008) 629–654, doi:10.1007/s00041-008-9035-z.

[142] P. Thiam, H.A. Kestler, F. Schwenker, Two-stream attention network for pain recognition from video sequences, Sensors (Switzerland) 20 (3) (2020) 839, doi:10.3390/s20030839.

[143] M. Yu, Y. Sun, B. Zhu, L. Zhu, Y. Lin, X. Tang, Y. Guo, G. Sun, M. Dong, Diverse frequency band-based convolutional neural networks for tonic cold pain assessment using EEG, Neurocomputing 378 (2020) 270–282, doi:10.1016/j.neucom.2019.10.023.

[144] J. Wang, M. Wei, L. Zhang, G. Huang, Z. Liang, L. Li, Z. Zhang, An autoencoder-based approach to predict subjective pain perception from high-density evoked EEG potentials, in: 2020 International Conference of the IEEE Engineering in Medicine Biology Society, Vol. 2020-July, Institute of Electrical and Electronics Engineers Inc., 2020, pp. 1507–1511, doi:10.1109/EMBC44109.2020.9176644.

[145] R.F. Rojas, J. Romero, J. Lopez-Aparicio, K.-L. Ou, Pain assessment based on fNIRS using Bi-LSTM RNNs, in: 10th International IEEE/EMBS Conference on Neural Engineering (NER), IEEE, 2021, pp. 399–402, doi:10.1109/NER49283.2021.9441384.

[146] H. Lim, B. Kim, G.J. Noh, S.K. Yoo, A deep neural network-based pain classifier using a photoplethysmography signal, Sensors (Switzerland) 19 (2) (2019), doi:10.3390/s19020384.

[147] B. Hu, C. Kim, X. Ning, X. Xu, Using a deep learning network to recognise low back pain in static standing, Ergonomics 61 (10) (2018) 1374–1381, doi:10.1080/00140139.2018.1481230.

[148] D. Mamontov, I. Polonskaia, A. Skorokhod, E. Semenkin, V. Kessler, F. Schwenker, Evolutionary algorithms for the design of neural network classifiers for the classification of pain intensity, Multimodal Pattern Recognition of Social Signals in Human-Computer-Interaction, LNAI Vol. 11377 (2019) 84–100, doi:10.1007/978-3-030-20984-1_8.

[149] C.Y. Chang, J.J. Li, Application of deep learning for recognizing infant cries, in: 2016 IEEE International Conference on Consumer Electronics-Taiwan, ICCE-TW 2016, Institute of Electrical and Electronics Engineers Inc., 2016, doi:10.1109/ICCE-TW.2016.7520947.

[150] M.S. Salekin, G. Zamzmi, R. Paul, D. Goldgof, R. Kasturi, T. Ho, Y. Sun, Harnessing the power of deep learning methods in healthcare: neonatal pain assessment from crying sound, in: 2019 IEEE Healthcare Innovations and Point of Care Technologies, HI-POCT 2019, Institute of Electrical and Electronics Engineers Inc., 2019, pp. 127–130, doi:10.1109/HI-POCT45284.2019.8962827.

[151] P. Thiam, F. Schwenker, Combining deep and hand-crafted features for audio-based pain intensity classification, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) LNAI, Vol. 11377, 2019, pp. 49–58, doi:10.1007/978-3-030-20984-1_5.

[152] F.S. Tsai, Y.M. Weng, C.J. Ng, C.C. Lee, Embedding stacked bottleneck vocal features in a LSTM architecture for automatic pain level classification during emergency triage, in: 2017 7th International Conference on Affective Computing and Intelligent Interaction, ACII 2017, Vol. 2018-Janua, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 313–318, doi:10.1109/ACII.2017.8273618.

[153] M.S. Salekin, G. Zamzmi, D. Goldgof, R. Kasturi, T. Ho, Y. Sun, Multi-channel neural network for assessing neonatal pain from videos, in: IEEE International Conference on Systems, Man and Cybernetics, Vol. 2019-Octob, Institute of Electrical and Electronics Engineers Inc., 2019, pp. 1551–1556, doi:10.1109/SMC.2019.8914537.

[154] E. Kasaeyan Naeini, S. Shahhosseini, A. Subramanian, T. Yin, A.M. Rahmani, N. Dutt, An edge-assisted and smart system for real-time pain monitoring, in: Proceedings - 4th IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies, CHASE 2019, Institute of Electrical and Electronics Engineers Inc., 2019, pp. 47–52, doi:10.1109/CHASE48038.2019.00023.

[155] A. Mauricio, J. Peña, E. Dianderas, L. Mauricio, J. Díaz, A. Morán, Chronic pain estimation through deep facial descriptors analysis, Commun. Comput. Inform. Sci. 1070 (2020) 173–185, doi:10.1007/978-3-030-46140-9_17.

[156] J. Ting, Y.-C. Yang, L.-C. Fu, C.-L. Tsai, C.-H. Huang, Distance ordering: a deep supervised metric learning for pain intensity estimation, in: 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), 2021, pp. 1083–1088, doi:10.1109/ICMLA52953.2021.00177.

[157] P. Thiam, P. Bellmann, H.A. Kestler, F. Schwenker, Exploring deep physiological models for nociceptive pain recognition, Sensors 19 (20) (2019) 4503, doi:10.3390/s19204503.

[158] A. Al-Qerem, An efficient machine-learning model based on data augmentation for pain intensity recognition, Egypt. Inform. J. 21 (4) (2020) 241–257, doi:10.1016/j.eij.2020.02.006.

[159] R. Zhi, C. Zhou, J. Yu, T. Li, G. Zamzmi, Multimodal-based stream integrated neural networks for pain assessment, IEICE Trans. Inf. Syst. E104D (12) (2021) 2184–2194, doi:10.1587/transinf.2021EDP7065.

[160] M.S. Salekin, G. Zamzmi, D. Goldgof, R. Kasturi, T. Ho, Y. Sun, Multimodal spatio-temporal deep learning approach for neonatal postoperative pain assessment, Comput. Biol. Med. 129 (2021), doi:10.1016/j.compbiomed.2020.104150.

[161] R. Wang, K. Xu, H. Feng, W. Chen, Hybrid RNN-ANN based deep physiological network for pain recognition, in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, Vol. 2020-July, Institute of Electrical and Electronics Engineers Inc., 2020, pp. 5584–5587, doi:10.1109/EMBC44109.2020.9175247.

[162] P. Thiam, H.A. Kestler, F. Schwenker, Multimodal deep denoising convolutional autoencoders for pain intensity classification based on physiological signals, in: ICPRAM 2020 - Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods, SCITEPRESS - Science and Technology Publications, 2020, pp. 289–296, doi:10.5220/0008896102890296.

[163] P. Thiam, H. Hihn, D.A. Braun, H.A. Kestler, F. Schwenker, Multi-modal pain intensity assessment based on physiological signals: a deep learning perspective, Front. Physiol. 12 (2021), doi:10.3389/fphys.2021.720464.

[164] S.D. Subramaniam, B. Dass, Automated nociceptive pain assessment using physiological signals and a hybrid deep learning network, IEEE Sens. J. 21 (3) (2021) 3335–3343, doi:10.1109/JSEN.2020.3023656.

[165] Y. Zhao, F. Ly, Q. Hong, Z. Cheng, T. Santander, H.T. Yang, P.K. Hansma, L. Petzold, How Much Does It Hurt: A Deep Learning Framework for Chronic Pain Score Assessment, in: C.A.Z.C.W.X., G. Di Fatta, V. Sheng (Eds.), IEEE International Conference on Data Mining Workshops, ICDMW, Vol. 2020-Novem, IEEE Computer Society, 2020, pp. 651–660, doi:10.1109/ICDMW51313.2020.00092.

[166] X. Yuan, M. Mahmoud, ALANet: Autoencoder-LSTM for pain and protective behaviour detection, in: Proceedings - 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2020, 2020, pp. 824–828, doi:10.1109/FG47880.2020.00063.

[167] Y. Li, S. Ghosh, J. Joshi, S. Oviatt, LSTM-DNN based approach for pain intensity and protective behaviour prediction, in: G.-F. F. Struc V. (Ed.), Proceedings - 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2020, Institute of Electrical and Electronics Engineers Inc., 2020, pp. 819–823, doi:10.1109/FG47880.2020.00061.

[168] P. Gouverneur, F. Li, W.M. Adamczyk, T.M. Szikszay, K. Luedtke, M. Grzegorzek, Comparison of feature extraction methods for physiological signals for heat-based pain recognition, Sensors 21 (14) (2021), doi:10.3390/s21144838.

[169] Y. Li, S. Ghosh, J. Joshi, PLAAN: pain level assessment with anomaly-detection based network, J. Multimodal User Interfaces (2021), doi:10.1007/s12193-020-00362-8.

[170] T. Lin, Y. Wang, X. Liu, X. Qiu, A survey of transformers, AI Open 3 (2022) 111–132, doi:10.1016/j.aiopen.2022.10.001.

[171] P. Werner, A. Al-Hamadi, K. Limbrecht-Ecklundt, S. Walter, S. Gruss, H.C. Traue, Automatic pain assessment with facial activity descriptors, IEEE Trans. Affect. Comput. 8 (3) (2017) 286–299, doi:10.1109/TAFFC.2016.2537327.

[172] K. M, L. S, The faces of pain: a cluster analysis of individual differences in facial activity patterns of pain, Eur. J. Pain. 18 (6) (2014) 813–823, doi:10.1002/J.1532-2149.2013.00421.X.

[173] M. Ranger, C. Johnston, K. Anand, Current controversies regarding pain assessment in neonates, Semin. Perinatol. 31 (2007) 283–288, doi:10.1053/j.semperi.2007.07.003.

[174] K.E. Boerner, K.A. Birnie, L. Caes, M. Schinkel, C.T. Chambers, Sex differences in experimental pain among healthy children: a systematic review and meta-analysis, Pain 155 (5) (2014) 983–993, doi:10.1016/j.pain.2014.01.031.

[175] S. Gkikas, C. Chatzaki, E. Pavlidou, F. Verigou, K. Kalkanis, M. Tsiknakis, Automatic pain intensity estimation based on electrocardiogram and demographic factors, in: Proceedings of the 8th International Conference on Information and Communication Technologies for Ageing Well and e-Health - ICT4AWE, INSTICC, SciTePress, 2022, pp. 155–162, doi:10.5220/0010971700003188.

[176] M.R. Jones, K.P. Ehrhardt, J.G. Ripoll, B. Sharma, I.W. Padnos, R.J. Kaye, A.D. Kaye, Pain in the elderly, 2016. 10.1007/s11916-016-0551-2

[177] R. Ochi, A. Midorikawa, Decline in emotional face recognition among elderly people may reflect mild cognitive impairment, Front. Psychol. 12 (2021), doi:10.3389/fpsyg.2021.664367.

[178] L.P. Forsythe, B. Thorn, M. Day, G. Shelby, Race and sex differences in primary appraisals, catastrophizing, and experimental pain outcomes, J. Pain 12 (5) (2011) 563–572, doi:10.1016/J.JPAIN.2010.11.003.

[179] P. Tschandl, N. Codella, B.N. Akay, G. Argenziano, R.P. Braun, H. Cabo, D. Gutman, A. Halpern, B. Helba, R. Hofmann-Wellenhof, A. Lallas, J. Lapins, C. Longo, J. Malvehy, M.A. Marchetti, A. Marghoob, S. Menzies, A. Oakley, J. Paoli, S. Puig, C. Rinner, C. Rosendahl, A. Scope, C. Sinz, H.P. Soyer, L. Thomas, I. Zalaudek, H. Kittler, Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study, Lancet Oncol. 20 (7) (2019) 938–947, doi:10.1016/S1470-2045(19)30333-X.

[180] G. Yang, Q. Ye, J. Xia, Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: a mini-review, two showcases and beyond, Inform. Fusion 77 (2022) 29–52, doi:10.1016/j.inffus.2021.07.016.

[181] K. Lekadir, R. Osuala, C. Gallin, N. Lazrak, K. Kushibar, G. Tsakou, S. Aussó, L.C. Alberich, K. Marias, M. Tsiknakis, S. Colantonio, N. Papanikolaou, Z. Salahuddin, H.C. Woodruff, P. Lambin, L. Martí-Bonmatí, FUTURE-AI: guiding principles and consensus recommendations for trustworthy artificial intelligence in medical imaging, 2021. 10.48550/arxiv.2109.09658

[182] P. Linardatos, V. Papastefanopoulos, S. Kotsiantis, Explainable AI: a review of machine learning interpretability methods, Entropy 23 (1) (2021) 1–45, doi:10.3390/e23010018.

[183] W.J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, B. Yu, Definitions, methods,

and applications in interpretable machine learning, Proc. Natl. Acad. Sci. 116 (44) (2019) 22071–22080, doi:10.1073/pnas.1900654116.

[184] R.E. Jack, Culture and facial expressions of emotion, Vis. Cogn. 21 (9–10) (2013) 1248–1286, doi:10.1080/13506285.2013.835367.

[185] L.E. McClelland, J.A. McCubbin, Social influence and pain response in women and men, J. Behav. Med. 31 (5) (2008) 413–420, doi:10.1007/s10865-008-9163-6.

[186] E. Cambria, N. Howard, J. Hsu, A. Hussain, Sentic blending: scalable multimodal fusion for the continuous interpretation of semantics and sentics, in: 2013 IEEE Symposium on Computational Intelligence for Human-like Intelligence (CIHLI), 2013, pp. 108–117, doi:10.1109/CIHLI.2013.6613272.