TABLE 1
Summary of datasets containing facial expressions of pain that are available upon request via email to first author or through a website.

Reference	Diagnostic Status	Pain Stimulus	Demographics	Sample Size	Annotation Granularity	Annotation (Labels)
Infant COPE [50] [55]	healthy	heel lancing	26 neonates (18–72 hours); 13 male, 13 female; Caucasian	204 facial images (pain images: 60)	frame-level	pain, crying, heel friction, nasal air stimulus, rest
UNBC-McMaster Shoulder Pain Expression Archive Database [18]	shoulder pain	range of motion tests on shoulders	129 adults; 63 male, 66 female	200 image sequences (total frames: 48,398; pain frames: 8369)	frame-level	12 AUs and their intensities (A–E), 66 facial landmarks, PSPI score
					sequence-level	self-report via VAS, sensory scale, affective-motivational scale; observer report via Observer Rated Pain Intensity (OPI)
BioVid Heat Pain Database [52]	healthy	heat	Part A: 87 adults (18–65 years); 44 male, 43 female	8700 videos (pain videos: 6960)	sequence-level	baseline (no pain), 4 pain stimulus intensity levels
			Part B: 86 adults (18–65 years); 42 male, 44 female	8600 videos (pain videos: 6880; partial facial occlusion due to facial EMG electrodes)	sequence-level	baseline (no pain), 4 pain stimulus intensity levels
			Part C: 87 adults (18–65 years); 44 male, 43 female	87 videos (long version of Part A with one video per subject)	segment-level	pain stimulus
		case vignette	Part D: 90 adults (18–65 years); 45 male, 45 female	630 videos (posed pain videos: 90)	sequence-level	7 posed expressions: neutral, pain, anger, disgust, fear, happy, sad
Hi4D-ADSIP [68]	healthy	none	80 adults (18–60 years); 32 male, 48 female; diverse ethnicities	3360 3D sequences (pain sequences: 240)	sequence-level	posed pain, 6 posed emotions (anger, disgust, surprise, fear, sadness, happiness), and 7 other facial articulations at 3 intensity levels each: mild, normal, extreme
BP4D-Spontaneous [69]	healthy	cold	41 adults (18–29 years); 18 male, 23 female; 20 Euro-American, 11 Asian, 6 African-American, 4 Hispanic	328 2D and 328 3D videos (pain videos: 41 2D, 41 3D)	frame-level	27 AUs (max. 20 sec segments), 3D head rotation angles, facial landmarks (83 for 3D, 49 for 2D)
					sequence-level	pain, anger, startle, fear, sadness, disgust, embarrassment, happiness

Spatial features consist of *geometric* or *textural* features. Geometric features describe the shape of the face in terms of point-based shape description schemes. These define point placements on facial features such as eyes, eyebrows, cheek, nose, lips, chin, and/or facial boundary. The locations of these facial feature points or higher-order features such as distances and angles between the facial feature points, are used as geometric features. Textural features describe the appearance of the face and facial features. Textural features include a description of the edges of facial features, and the wrinkles or folds that appear on or around them. Textural features used in literature range from raw pixel intensities to hand-crafted or self-learned features. The commonly used hand-crafted textural feature descriptors are Gabor filters [133], Local Binary Patterns (LBP) [134] [135], and Histogram of Oriented Gradients (HOG) [136]. Geometric features were rarely used alone. Textural features, either alone or in combination with geometric features, are the most widely used features in automatic pain detection. The combination of geometric and textural features are denoted in Tables 4 and 5 as *hybrid* features.

Spatiotemporal features describe changes in spatial features over time, and can be categorized in a similar fashion as spatial features into geometric, textural, and hybrid categories. In other words, spatiotemporal geometric and textural features were used either independently or in combination (hybrid). Geometric features extracted from a sequence of images were summarized using mathematical and statistical operators. Spatiotemporal textural features such as LBP-TOP [137] and HOG-TOP [138] were extracted from Three Orthogonal Planes (TOP), one of which covers the temporal dimension that spans a temporally ordered sequence of images. Yang et al. [104] compared the performance of several spatiotemporal textural features, such as LBP-TOP, LPQ-TOP[11], BSIF-TOP[12], and their combinations.

In cases where hybrid, mixed, or multiple features of the same type were used, the fusion of features was performed either before the learning step (cf. [75], [123]), or by fusing the decisions of classifiers trained separately for each feature (cf. [107], [116]). The former is commonly referred to as "early fusion", and the latter is commonly referred to as "late fusion".

In two-step approaches, pain detection is done based on an intermediate representation of the face in terms of

11. LPQ-TOP is the spatiotemporal variant of Local Phase Quantization (LPQ) [139]
12. BSIF-TOP is the spatiotemporal variant of Binarized Statistical Image Features (BSIF) [140]

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2019.2958341, IEEE Transactions on Pattern Analysis and Machine Intelligence

TABLE 2
Summary of datasets whose availability is unknown at the time of writing this survey. The authors of the respective papers might be contacted for potential access. Note: 'elderly' denotes adults aged over 65 years.

Reference	**Diagnostic Status**	**Pain Stimulus**	**Demographics**	**Sample Size**	**Annotation Granularity**	**Annotation (Labels)**
Wilkie [58]	lung cancer	Activities of Daily Living (ADL)	43 adults; 27 male, 16 female; 31 Caucasian, 12 others	43 videos	segment-level	9 AUs (20 sec segments)
					sequence-level	self-report via VAS and State-Trait Anxiety Inventory (STAI)
Roy et al. [70]	healthy	none	34 adults	1088 videos (pain videos: 136)	sequence-level	pain, neutral, 6 basic emotions (anger, sadness, happiness, disgust, fear, surprise)
Kunz et al. [53]	demented	mechanical pressure	42 elderly (mean: 76.7 years); 20 male, 22 female	42 videos (pain stimulus sessions: 840)	segment-level	for each 5 sec stimulus session: 44 AUs and their intensities (A–E), self-report of pain level via verbal category scale
	healthy		54 elderly (mean: 74.2 years); 11 male, 43 female	54 videos (pain stimulus sessions: 1080)		
Lu et al. [51]	healthy	heel lancing	57 neonates; 30 male, 27 female	510 images (pain images: 160)	frame-level	pain, cry, calm
Hammal et al. [71]	healthy	heat	20 adults	20 videos (pain stimulus sessions: 40)	segment-level	for each 5 sec stimulus session: 44 AUs and their intensities (A–E)
Kunz et al. [57]	healthy	heat	44 young adults (18–30 years); 22 male, 22 female	44 videos (pain stimulus sessions: 352)	segment-level	for each 5 sec segment after stimulus reached peak: 44 AUs and their intensities (A–E), self-report via VAS
Littlewort et al. [56] [72]	healthy	cold	26 adults; 6 male, 20 female	78 one-minute videos (real pain videos: 26; faked pain videos: 26)	sequence-level	baseline (no pain), real pain, faked pain
Niese et al. [54]	healthy	hand movements with tourniquet attached	21 adults (20–30 years); 10 male, 11 female	21 image sequences (total frames: 966000; pain frames: 31500)	segment-level	self-report of pain intensity via NRS
EmoPain [27]	chronic lower back pain	physical exercises	22 adults (19–67 years); 7 male, 15 female; 18 Caucasian, 4 others	44 videos (total frames: 585,487; pain frames: 50,071)	frame-level	pain, no pain
					segment-level	self-report of pain and anxiety on 1–10 scale
	healthy		28 adults (mean age: 37.1 years); 14 male, 14 female; 26 Caucasian, 2 Asian	–	–	no pain
Irani et al. [66]	healthy	mechanical pressure	12 elderly females (66–90 years)	96 videos (total frames: 2388; pain frames: 1631)	sequence-level	self-report of pain intensity via NRS
Pediatric Pain Dataset [59] [73]	after appendectomy	endogenous and exogenous (manual pressure at surgical site)	50 youth (5–18 years); 27 male, 23 female; 35 Hispanic, 9 non-Hispanic white, 5 Asian, 1 Native American	300 videos (endogenous pain: 150 exogenous pain: 150)	sequence-level	self and observer reports of pain intensity via NRS
Singh [60]	back/neck/knee pain	manual pressure on affected area	21 adults; 12 male, 9 female	21 image sequences (total frames: 336)	frame-level	7 AUs and their intensities
Tsai et al. [28]	emergency cases with pain or headache	endogenous	117 adults	205 videos	sequence-level	self-report via NRS

TABLE 3
Summary of the learning approaches that have been developed and tested for automatic pain detection from facial expressions.

Learning Task	Temporal Information	References
One-Step Approaches		
pain and no-pain	no	Brahnam et al. [79], Monwar and Rezaei [80], Brahnam et al. [81], Lu et al. [51], Ashraf et al. [82], Lucey et al. [83], Siebers et al. [84], Nanni et al. [85], Gholami et al. [86], Monwar and Rezaei [87], Wei and Li-min [88], Lucey et al. [18], Lucey et al. [89], Werner et al. [90], Chen et al. [91], Khan et al. [92], Pedersen [93], Neshov and Manolova [94], Rathee and Ganotra [95], Aung et al. [27], Kharghanian et al. [96], Roy et al. [97], Rupenga and Vadapalli [98], Meawad et al. [99], Alphonse and Dharma [100]
	yes	Werner et al. [101], Meng and Bianchi-Berthouze [102], Werner et al. [29], Kächele et al. [103], Yang et al. [104]
pain and emotions	no	Niese et al. [54]
	yes	Hammal et al. [71], Hammal and Kunz [105]
pain and states (crying, calm/rest)	no	Brahnam et al. [79], Lu et al. [51], Yuan et al. [106]
pain and distress (via heel friction or air stimulus on nose)	no	Brahnam et al. [79]
pain intensity (continuous)	no	Werner et al. [90], Kaltwang et al. [107], Romera-Paredes et al. [108], Neshov and Manolova [94], Wang et al. [109], Liu et al. [110]
	yes	Kächele et al. [103], Florea et al. [111], Zhou et al. [112], Kaltwang et al. [113], Zhao et al. [114], Rodriguez et al. [115], Egede et al. [116], Egede and Valstar [117], Lopez-Martinez et al. [118], Tavakolian and Hadid [119]
pain intensity (discrete)	no	Gholami et al. [86], Lucey et al. [89], Hammal and Cohn [120], Singh [60], Rathee and Ganotra [95], Roy et al. [97], Alphonse and Dharma [100]
	yes	Rudovic et al. [121], Irani et al. [122], Irani et al. [66], Werner et al. [123], Tsai et al. [28], Lopez-Martinez et al. [118]
pain event in sequence	yes	with localization: Sikka et al. [124], Sikka et al. [125], Lo Presti and La Cascia [126], Lo Presti and La Cascia [127]
		without localization: Chen et al. [75]
Two-Step Approaches		
pain and no-pain	no	Lucey et al. [83], Lucey et al. [128], Zafar and Khan [129]
	yes	Schmid et al. [77], Sikka et al. [59], Siebers et al. [78]
pain intensity (continuous)	yes	Sikka [73], Sikka et al. [59], Zhang et al. [76], Lopez-Martinez et al. [130]
pain intensity (discrete)	no	Zafar and Khan [129]
	yes	Ghasemi et al. [74]
posed and genuine pain	yes	Littlewort et al. [72], Littlewort et al. [56], Bartlett et al. [131]

AUs. Features used for learning pain-related targets were therefore extracted from the AU labels or AU scores[13] provided by the first step. We categorize these features that are indirect representations of the input image or image sequence into *non-temporal* and *temporal* features. Table 6 provides an overview of the indirect features that have been used for automatic pain detection. Non-temporal features refer to the AU representations for a single image or a *single timestep* in an image sequence. In this case, the AU labels or scores for the image are used as features for pain detection (e.g. [77], [78], [83]). Temporal features refer to AU representations for a sequence of images spanning *multiple timesteps*. In this case, AU scores provided by the first learning stage are aggregated using statistical operators (cf. [59]) or dynamic features are extracted using temporal filters (cf. [131]). Note that the categorization into non-temporal and temporal features is based purely on whether the pain detection in the second step used AU detection outputs for a single image/timestep or for multiple timesteps. It does not take into account whether temporal information was used in the first step for AU detection. It was noted that the two-step approach followed by Lopez-Martinez et al. [130] used a combination of direct and indirect features for continuous pain intensity estimation (see Table 6).

The extracted features are often post-processed to increase their discriminative power or to extract the most important information. Principal Component Analysis (PCA) is a commonly used method to select the most important feature dimensions and thereby transform the features into a lower-dimensional space (cf. [76], [79], [114]). Rathee and Ganotra [95] proposed multiview distance metric learning to fuse LBP, HOG, and Gabor features, and to increase the discriminative power of the new set of features. Florea et al. [111] used a semi-supervised transfer learning method based on spectral regression to learn the most discriminative feature dimensions of the extracted Histogram of Topological (HoT) features and to reduce the dimensionality of the feature space. An exhaustive survey of the feature post-processing methods is outside the scope of this paper. The reader is advised to refer to other surveys on facial expression analysis (e.g. [132]) to obtain an overview about

13. The term “scores” is used in this paper to broadly refer to scores/probabilities/intensities of AUs.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2019.2958341, IEEE Transactions on Pattern Analysis and Machine Intelligence

TABLE 4
Summary of spatial representations extracted directly from facial images for automatic pain detection.

Feature Sub-Type	Features	References
geometric	facial landmark positions	Meng and Bianchi-Berthouze [102], Ghasemi et al. [74], Aung et al. [27], Rupenga and Vadapalli [98], Liu et al. [110], Lopez-Martinez et al. [118]
	facial landmark distances	Romera-Paredes et al. [108], Meawad et al. [99]
	facial landmark distances and angles	Niese et al. [54], Siebers et al. [84]
	facial landmark positions, distances, angles	Zafar and Khan [129]
textural	pixel intensities	Brahnam et al. [79], Gholami et al. [86], Ghasemi et al. [74]
	Gabor filters	Littlewort et al. [72], Lu et al. [51], Yuan et al. [106], Littlewort et al. [56], Sikka [73], Bartlett et al. [131], Sikka et al. [59], Roy et al. [97]
	Discrete Cosine Transform (DCT)	Brahnam et al. [81], Aung et al. [27]
	Local Binary Pattern (LBP) or its variant	Nanni et al. [85], Chen et al. [91], Rudovic et al. [121], Aung et al. [27]
	Local Ternary Pattern (LTP) or its variant	Nanni et al. [85]
	histogram of quantized edge directions	Monwar and Rezaei [80]
	Histogram of Oriented Gradients (HOG) around facial landmarks	Chen et al. [75]
	Histogram of Topological (HoT) features	Florea et al. [111]
	variants of Local Directional Pattern (LDP)	Alphonse and Dharma [100]
	Scale-Invariant Feature Transform (SIFT)	Neshov and Manolova [94], Singh [60]
	Speeded-Up Robust Features (SURF)	Singh [60]
	pyramid HOG and pyramid LBP	Khan et al. [92]
	supervised locality preserving projection	Wei and Li-min [88]
	log-normal filters	Hammal and Cohn [120]
	Gabor filters, HOG, and LBP	Rathee and Ganotra [95]
	3D binary edges	Zhang et al. [76]
	features learned by semi-supervised autoencoder	Pedersen [93]
	deep learned features	Kharghanian et al. [96], Rodriguez et al. [115], Wang et al. [109]
hybrid (geometric + textural)	facial landmark distances, nasal root wrinkles, context variable	Hammal et al. [71], Hammal and Kunz [105]
	facial landmark distances, histogram of quantized edge directions	Monwar and Rezaei [87]
	facial landmark positions, LBP, Gabor filters	Zhao et al. [114]
	facial landmark distances, mean gradient magnitude in facial regions	Werner et al. [90]
	facial landmark positions, DCT, LBP	Kaltwang et al. [107]
	facial landmark positions, pixel intensities	Ashraf et al. [82], Lucey et al. [128], Lucey et al. [18], Lucey et al. [89]
	facial landmark positions, DCT	Lucey et al. [83]
	facial landmark positions, distances, angles, HOG	Egede et al. [116]

commonly used feature post-processing or feature selection methods.

5.3 Learning Methods

The different learning tasks examined by the existing automatic pain detection approaches were discussed in Section 5.1 and listed in Table 3. The majority of the learning tasks were either binary or multiclass classification tasks such as pain versus no-pain, genuine versus posed pain, discrete pain intensity levels, and pain versus emotions. The other type of learning tasks were regression tasks for estimating pain intensity as a continuous-valued function. Table 7 lists the machine learning methods that have been used in the reviewed literature for pain-related classification and regression tasks. Two-step approaches involve two learning tasks: AU detection and pain detection. AU detection tasks include binary or multiclass classification for detecting the presence of different AUs, and regression for estimating intensities of different AUs. The machine learning methods used for AU detection in two-step approaches are also listed in Table 7. Certain one-step approaches did not use machine learning methods for pain detection. Irani et al. [122] [66] used experimentally determined thresholds on spatiotemporal features to determine three discrete levels of pain. Meawad et al [99] defined a mapping between facial landmark distances and pain-related AUs. Based on this mapping, the PSPI scale was modified. Sequence-level pain detection was then performed by checking whether a predefined number of consecutive frames showed the presence of pain according to the modified PSPI scale. Certain two-step approaches did not use machine learning methods in the second step. Zafar and Khan [129] applied the PSPI scale on the discrete AU intensities predicted by a set of k-nearest neighbor classifiers. Zhang et al. [76] averaged the probabilities of selected pain-related AUs to calculate the pain intensity estimate.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2019.2958341, IEEE Transactions on Pattern Analysis and Machine Intelligence

TABLE 5
Summary of spatiotemporal and mixed representations extracted directly from facial images for automatic pain detection.

Feature Sub-Type	Features	References
Spatiotemporal Features		
geometric	Hankel matrices based on facial landmark positions and/or distances	Lo Presti and La Cascia [126]
	statistical features from sequence of facial landmark distances and quadratic polynomial coefficients of mouth shape	Tsai et al. [28]
	bag of words from k-means based clusters of sequence of geometric features (facial landmark distances and quadratic polynomial coefficients of mouth shape)	Tsai et al. [28]
textural	HOG from Three Orthogonal Planes (HOG-TOP)	Chen et al. [75]
	LBP from Three Orthogonal Planes (LBP-TOP)	Kaltwang et al. [113]
	combinations of LBP-TOP, LPQ-TOP, BSIF-TOP	Yang et al. [104]
	energy from optical flow	Ghasemi et al. [74]
	time-integral of histogram of oriented energies	Irani et al. [122], Irani et al. [66]
	Hankel matrices from time series of Haar and/or Gabor features	Lo Presti and La Cascia [127]
	deep learned spatiotemporal features	Zhou et al. [112], Egede et al. [116], Tavakolian and Hadid [119]
	max temporal pooling on sequence of SIFT based features	Sikka et al. [124], Sikka et al. [125]
hybrid (geometric + textural)	statistical features from sequence of head pose, facial landmark distances, mean gradient magnitudes	Werner et al. [101], Werner et al. [29]
	statistical and time features from sequence of head pose, facial landmark distances, mean gradient magnitudes	Werner et al. [123]
	LBP-TOP, statistical features from facial distances	Kächele et al. [103]
Mixed (Spatial + Spatiotemporal) Features		
texture	HOG around facial landmarks, HOG-TOP	Chen et al. [75]
hybrid (geometric + textural)	facial landmark positions, distances, angles, HOG, deep learned features from image sequence	Egede et al. [116], Egede and Valstar [117]

TABLE 6
Summary of indirect representations of facial images and image sequences used in two-step automatic pain detection approaches either alone or in combination with direct representations.

Feature Category	Features	References
Indirect Features		
non-temporal or single timestep	AU scores	Lucey et al. [83], Lucey et al. [128], Zafar and Khan [129], Zhang et al. [76]
	AU labels	Schmid et al. [77], Siebers et al. [78]
temporal or multiple timesteps	statistical features from AU scores	Littlewort et al. [72], Sikka [73], Sikka et al. [59]
	temporal filters on AU scores	Littlewort et al. [56], Bartlett et al. [131]
	histogram of AUs	Ghasemi et al. [74]
Combined (Direct + Indirect) Features		
temporal or multiple timesteps	statistical features from sequence of AU scores, facial landmark distances, eye gaze coordinates, and head pose	Lopez-Martinez et al. [130]

Almost all classification and regression tasks were supervised. Ground truth in the form of pain or AU labels, and discrete or continuous-valued pain or AU intensities, were used to train the machine learning models. Support Vector Machines (SVM) and its variants such as multiple kernel SVM and Support Vector Regressors (SVR), are the most widely used supervised machine learning methods. Probabilistic methods such as Relevance Vector Machines (RVM) and different variants of conditional random fields, have been used for supervised classification tasks. Random forests have been used for both supervised classification and supervised regression tasks in automatic pain detection. Less commonly used methods for supervised learning include decision trees and classical regression methods such as linear and logistic regression. More recently, deep learning methods such as Convolutional Neural Networks (CNN), Recurrent CNN, and Long Short-Term Memory (LSTM) recurrent neural networks, are increasingly being used for end-to-end learning of pain intensities from single images (e.g. [109]) or image sequences (e.g. [115], [112]).

Very few works explored machine learning strategies other than supervised learning. For example, an unsupervised comparative learning method was used by Werner et al. in [90] for estimating continuous-valued pain intensity; Sikka et al. [124], [125] employed weakly supervised learning with the help of a multiple-instance variant of boosting algorithms for pain event localization in an image sequence. Semi-supervised learning strategies have not yet been explored in the context of automatic pain detection from facial expressions.

The metrics used to quantify the performance of automatic pain detection from facial expressions depend on the learning task. For classification tasks, metrics such as accuracy, F1 score, and area under Receiver Operating

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2019.2958341, IEEE Transactions on Pattern Analysis and Machine Intelligence

TABLE 7
Summary of machine learning methods used in the automatic pain detection approaches.

Prediction Task	Approach	Machine Learning Method	References
Supervised Methods			
classification	one-step	Support Vector Machine (SVM)	Brahnam et al. [79], Brahnam et al. [81], Lu et al. [51], Monwar and Rezaei [87], Lucey et al. [83], Ashraf et al. [82], Niese et al. [54], Siebers et al. [84], Nanni et al. [85], Gholami et al. [86], Lucey et al. [18], Lucey et al. [89], Hammal and Cohn [120], Werner et al. [90], Werner et al. [101], Khan et al. [92], Pedersen [93], Neshov and Manolova [94], Lo Presti and La Cascia [126], Singh [60], Chen et al. [75], Aung et al. [27], Rathee and Ganotra [95], Kharghanian et al. [96], Roy et al. [97], Werner et al. [123], Yang et al. [104], Rupenga and Vadapalli [98], Tsai et al. [28]
		Relevance Vector Machine (RVM)	Gholami et al. [86]
		random forest	Khan et al. [92], Werner et al. [29], Kächele et al. [103], Werner et al. [123]
		multiple kernel SVM	Wei and Li-Min [88], Chen et al. [75]
		Neural Network (NN)	Monwar and Rezaei [80]
		Neural Network Simultaneous Optimization Algorithm (NNSOA)	Brahnam et al. [81]
		extreme learning machine	Rupenga and Vadapalli [98], Alphonse and Dharma [100]
		decision tree	Khan et al. [92]
		k-nearest neighbors	Siebers et al. [84], Khan et al. [92], Lo Presti and La Cascia [126]
		k-nearest neighbors + hidden Markov model	Meng and Bianchi-Berthouze [102]
		Adaboost or its variants	Yuan et al. [106], Chen et al. [91], Lo Presti and La Cascia [127]
		transferable belief model	Hammal et al. [71], Hammal and Kunz [105]
		heteroscedastic conditional ordinal random field	Rudovic et al. [121]
		hidden conditional random field	Lopez-Martinez et al. [118]
		regularized multi-task learning	Romera-Paredes et al. [108]
	two-step	SVM	step1-AU: Lucey et al. [83], Lucey et al. [128] step2-pain: Bartlett et al. [131] both steps: Littlewort et al. [72], Littlewort et al. [56], Ghasemi et al. [74]
		logistical linear regression	step2-pain: Lucey et al. [83], Lucey et al. [128]
		k-nearest neighbors	step1-AU: Zafar and Khan [129]
		logistic regression	step2-pain: Sikka et al. [59]
		alignment-based learning	step2-pain: Schmid et al. [77], Siebers et al. [78]
		hidden conditional random field	step2-pain: Ghasemi et al. [74]
		latent-dynamic conditional random field	step1-AU: Zhang et al. [76]
regression	one-step	support vector regression	Florea et al. [111], Lopez-Martinez et al. [118]
		ordinal support vector regression	Zhao et al. [114]
		relevance vector regression or its variants	Kaltwang et al. [107], Kaltwang et al. [113], Egede et al. [116], Egede and Valstar [117]
		random forest	Kächele et al. [103]
		linear regression	Neshov and Manolova [94]
		ordinal support vector regression	Zhao et al. [114]
		NN	Lopez-Martinez et al. [118]
		Convolutional Neural Network (CNN)	Wang et al. [109]
		3D CNN with kernels of varying temporal lengths	Tavakolian and Hadid [119]
		recurrent CNN	Zhou et al. [112]
		LSTM recurrent neural network	Rodriguez et al. [115], Lopez-Martinez et al. [118]
	two-step	support vector regression	step1-AU: Bartlett et al. [131], Sikka et al. [59] both steps: Sikka [73]
		linear regression	step2-pain: Sikka et al. [59]
		multi-task NN	step2-pain: Lopez-Martinez et al. [130]
Weakly-Supervised Methods			
classification	one-step	multiple instance learning-Boost	Sikka et al. [124], Sikka et al. [125]
regression	one-step	ordinal support vector regression	Zhao et al. [114]
		NN + Gaussian process regression model	Liu et al. [110]
Unsupervised Methods			
regression	one-step	comparative learning	Werner et al. [90]
		ordinal support vector regression	Zhao et al. [114]