Survey of Affect Representation Schemes used in Vision-Based Automatic Affect Prediction

A Systematic Literature Review

Lucia Serrano Ruber[1]

Supervisor(s): Chirag Raman[1]**, Bernd Dudzik**[1]

[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 25, 2023

Name of the student: Lucia Serrano Ruber
Final project course: CSE3000 Research Project
Thesis committee: Chirag Raman, Bernd Dudzik, Cynthia Liem

Abstract

In human-human interactions, the majority of information is conveyed through body language, specifically facial expressions. Consequently, researchers have been interested in improving human-computer interactions through developing systems with automatic understanding of body language and facial expressions. This technology is especially useful due to its broad range of applications in fields such as healthcare, education, and safety & security. Vision-based automatic affect recognition (AAR) systems aim to predict a subject's affective state based on visual input such as image or video. These systems analyze and classify subjects' facial expressions and body language using affect representation schemes (ARS), most often classified as either categorical or dimensional. This paper explores the current state of ARS used in vision-based AAR through a systematic literature review following PRISMA guidelines. We selected 53 papers from WebOfScience according to our eligibility criteria which included computer science papers written in English proposing a vision-based AAR system targeting single subjects, and excluded studies dealing exclusively with micro-expressions or group affect recognition. Additionally, given the time limitation imposed on this research we excluded papers that were not readily accessible with our TU Delft license, used multimodal input, or did not use a dataset included in our predefined list. For this exploration we specifically look at the schemes used, the popularity and trends of usage, motivations, and psychological basis. From the 53 reviewed papers, all of the papers target utilitarian emotions using at least one discrete ARS. The most commonly used schemes classify affective states into *happiness, sadness, fear, anger, surprise,* and *disgust*. While the majority of papers are lacking in providing explicit reasoning for their choices, most ARS are based grounded in psychological theories. Our results show an established norm within this area of research. However, they also evidence a lack of displayed critical thought in the selection of schemes. This oversight limits potential for future AAR research.

1 Introduction

Automatic Affect Recognition (AAR)—or affect computing—aims to recognize the affective state of a subject by analysing one's physiological signals, speech, or body language, for instance. An affective state can refer to for example one's mood, attitude, or emotion [1]. This technology is becoming increasingly popular in the field of human-computer interaction. Its applications range from entertainment to healthcare to education to vehicle safety mechanisms, such as detecting driver drowsiness [2]–[4]. This makes it a topic of interest for many, even outside of computer science.

Vision-based AAR systems aim to predict a subject's affective state based on visual input such as image or video. Most of the information conveyed in human-human interactions comes from their facial expressions (55%), when compared to speech (38%) and language (7%) [5], emphasizing just how crucial the understanding of facial expressions is for effective communication. This is one of the reasons vision-based AAR is popular among affect computing research. Additionally, there are many datasets available for researchers to use for vision-based AAR (Cohn-Kanade (CK+) [6], Japanese Female Facial Expression (JAFFE) [7], and Radboud Faces Database (RaFD) [8] to name a few), so it is not necessary for them to manually create, acquire, and annotate lots of data which is time and resource intensive.

The COVID-19 pandemic has accelerated research in vision-based AAR introducing both new challenges and applications. With the enforcement of face masks and remote interactions these systems have been forced to adapt and evolve. Before the pandemic, most vision-based AAR systems relied on the entirety of one's face. Face masks made the problem of partial occlusion more relevant, bringing more attention and interest to for example techniques relying on one's eyes to mitigate this occlusion [9]. Remote interactions have driven development of other vision-based AAR applications. One such example is an educational aid for teachers to more easily notice struggling students in online learning environments [2], which is much harder to recognize remotely than in person.

Recognizing facial expressions can give insights into one's affective state but it should be noted that it is not possible to truly recognize the affective state displayed by an individual. It is merely an interpretation attempt [4]. That is, accurately recognizing one's facial expression as "happy" does not mean the subject is truly feeling happy. This problem is not unique to visual input, the same is true when performing AAR using speech or physiological signals as input. One's facial expression, voice, or heart rate, is not exclusively influenced by one's affective state. There is always a chance that a subject can (sub)consciously control these, hiding their true affective state.

To classify a subject's affective state in AAR, the system needs an Affect Representation Scheme (ARS). There are various ARS used in the existing literature, most of which are grounded in psychological theories. The majority of these schemes can be grouped into two groups: (1) categorical or (2) dimensional. A categorical model is one that proposes distinct affective states—or categories. One such example is an ARS based on Ekman's Basic Emotion Theory (BET), describing 6 basic and universal emotions: *joy, anger, fear, sadness, surprise*, and *disgust* [10]–[12] shown in Figure 1a. On the other hand, dimensional models propose a multi-dimensional scale with which affective states can be described, such as Russel's valence-arousal model [13] shown in Figure 1b.

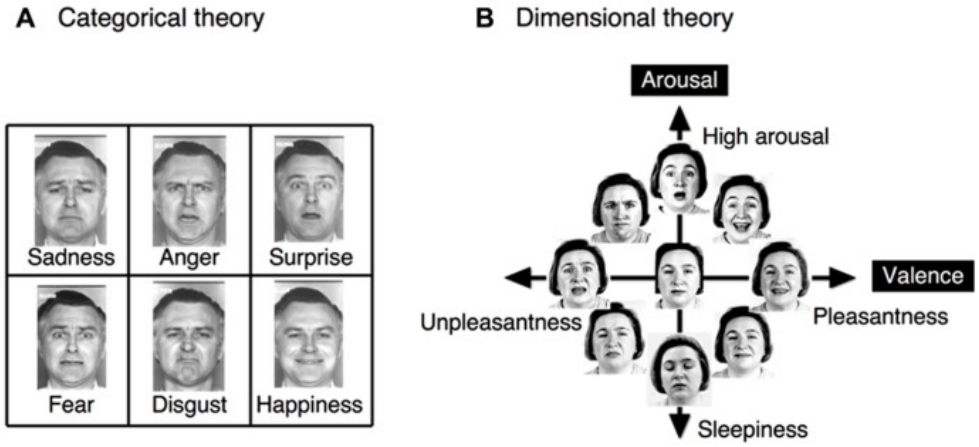


Figure 1: (A) An example of Ekman's model. (B) An example of a valence-arousal model [14].

Other reviews exploring the state of vision-based AAR have previously been conducted, but none with a focus on ARS. These other reviews mainly focus on the algorithms and approaches employed [15]–[19]. Our paper differs from this, offering a detailed look into the ARS used in these systems. It is important to have a clear understanding of ARS, how and when they are used, and the impact different ARS can have on AAR results. Having this information will allow for future researchers to make more informed decisions in the development of their AAR systems.

Our systematic literature review aims to provide these insights by exploring the research question: ***What is the current state of affect representation schemes used in automatic vision-based affect recognition systems?*** This has been broken down into the following sub-questions:

1. What types of affective states have been targeted by prediction systems?
2. What different affect representation schemes have been used for this, and if so, what is the motivation for this particular scheme?
3. Are systems using more than one emotion representation scheme simultaneously, and if so, what is their motivation for doing so?
4. What are the differences in the popularity of schemes used for modeling different affective states, if any?
5. How has the popularity of specific schemes changed over time, if at all?
6. Are the majority of representation schemes used based on psychological theory?

The report will take the following structure. Section 2 outlines our methodology and specific steps followed throughout the systematic literature review. Then, section 3 describes our results that we gathered. Section 4 provides a discussion and evaluation of the results. Section 5 outlines the relevant ethical considerations and their implications. We conclude in section 6 with remarks and recommendations for future work.

2 Methodology

This section describes the methodology employed throughout the review. 2.1 explains why we chose a systematic literature review for this research. 2.2 details the eligibility criteria we used to select the papers for this review. 2.3 motivates our choice of search engines used, and 2.4 outlines the search strategy implemented to obtain the final results. 2.5 and 2.6 detail the selection process and the search results obtained. Finally 2.7 briefly summarizes the data extraction process.

2.1 PRISMA

For this research, we chose to perform a systematic literature review following the PRISMA guidelines [20]. Systematic literature reviews, like other literature reviews, are performed to gather and synthesize large amounts of information on a certain topic. However, where other reviews might take a more general approach, a systematic review describes how the papers were collected and the search and filtering strategies employed. This emphasizes the structure with which the review is performed ensuring reproducibility of results.

2.2 Eligibility criteria

To systematically select papers to be reviewed, we need to establish a clear set of criteria to follow. 2.2.1 describes the general inclusion and exclusion criteria that were chosen in order to keep the papers reviewed relevant to the topic at hand. 2.2.2 introduces additional exclusion criteria that were added to keep the number of papers to review manageable within the given timeframe.

2.2.1 General eligibility criteria

Inclusion criteria This list contains the criteria to which a paper has to uphold in order to be deemed eligible for this study.

- **The paper proposes an AAR system.**
  This is the most basic filtering criteria as this is the general topic that is being reviewed.
- **The system described uses visual data as input.**
  This review is looking specifically into vision-based affect prediction, thus we want to review papers that tackle this problem.
- **The paper is in the *Computer Science* field.**
  There are many papers on affect prediction from other fields such as psychology and neuroscience that are not relevant for us in this review. To check this criterion, we use the labels given to the papers in the search databases.
- **The system described performs AAR on a single subject at a time.**
  To keep consistency in the review, we will only be reviewing papers whose proposed systems do not deal with group emotion recognition and instead only attempt to perform AAR on a single subject at a time.

Exclusion criteria If a paper conforms to any of these criteria, it is deemed ineligible for this study. These were established to ensure consistency and continuity between papers.

- **The paper is a survey or review paper analysing other papers.**
  These papers will not be introducing new methods but rather analysing existing approaches. We do not want to analyse the analyses.
- **The paper only introduces a new dataset to be used in future research, with no usage in an AAR system.**
  The dataset itself is not a novel method or approach for automatic affect recognition. It is not considered relevant for this research, unless the paper itself also proposes an AAR system using this data.
- **The paper is not written in English.**
  This could introduce some selection bias, but majority of papers in this scientific community are published in English so the effects of this bias are unlikely to have detrimental effects on this research.
- **The system described only deals with micro-expression recognition.**
  In this research we are interested in macro-expressions rather than micro-expressions — minor spontaneous muscle movements that can reveal underlying emotions. This is a different problem than what we are describing. Systems that deal with both macro- and micro-expressions are still considered eligible.
- **The system described deals with group affect recognition.**
  For this review we want to focus on single-subject affect recognition. Group affect recognition is a different problem and including this would broaden the scope of this research too much.

2.2.2 Feasibility constraints

These constraints were put in place to limit the number of papers to review to a manageable amount for the given time frame of 8 weeks.

- **The paper is not directly accessible with a TU Delft license or needs to be requested.** Due to time constraints, papers that are not immediately accessible to us will not be reviewed. The requesting and retrieval process can take extra time that we do not have. There are enough other papers to review, so excluding these only saves valuable time and resources.
- **The system described uses multiple modalities for input.** As mentioned in the inclusion criteria, we are interested in vision-based affect prediction. Vision-based does not necessarily mean vision-*only*. Looking into multimodal systems using for example audio or physiological signals to enhance performance, could give us a better understanding of the current state of research. However in the interest of time for this study, these are excluded and we are looking exclusively at AAR systems that only use visual input. If the system uses audio-visual input (such as a video clip), it must only use the image stream for it to be included in the review.
- **The study does not use at least one of the datasets listed in Appendix B.** We collected this list of datasets from other literature reviews on AAR [19], [21]. Excluding papers that do not use at least one of these data sets significantly reduces the amount of time needed to gather, synthesize, and analyse results.

2.3 Search Engines

While we considered multiple databases for this review, we ultimately decided on using only WebOfScience (WoS). Google Scholar has no easy way to export all the results from a query, making it tedious to work with. Additionally, Google Scholar retrieved over 3,000,000 results without the feasibility constraints described in 2.2. When we tried to limit the search results with our feasibility criteria, our search query exceeded Google Scholar's 256 character limit.

SCOPUS and WoS both proved more effective than Google Scholar, but in the interest of time, only the results from WoS were used. Both SCOPUS and WoS could export results, handle the search query with no troubles, and retrieve a more manageable amount of papers than Google Scholar. Additionally, SCOPUS and WoS both have useful filters at their disposal allowing for easy automatic first-pass exclusion of papers deemed ineligible according to the eligibility criteria described in 2.2, saving us valuable time and energy. Ideally we would have used both search engines, but within the given time constraints it was not realistic. Hence only WoS was used in the end as it retrieved a smaller amount of papers (549 compared to nearly 3,000 from SCOPUS).

2.4 Search Strategy

We employed the following search strategy. We built our search query in an iterative process. The initial query covered the main concepts of *vision*, *affect*, and *recognition*. From here, we added related terms found in relevant retrieved papers for a more representative collection of results. We did this several times with the updated keywords until we were satisfied with the retrieved results: not too many irrelevant results, but also not so niche where we might be missing relevant results. The chosen key words are shown in table 1. Additionally, (most) surveys and reviews were excluded directly through the query, automatically applying one of our exclusion criteria. Once the query was satisfactory, see Appendix A.1, it was expanded with relevant data sets to enforce the last exclusion criteria described in the feasibility constraints in 2.2. The query searched by *Topic* which searches "title, abstract, author keywords, and Keywords Plus", according to WoS. *Keywords Plus* are "words or phrases that frequently appear in the titles of an article's references, but do not appear in the title of the article itself" [22]. Along with this, we used WoS's built-in filters to limit the search to papers in the Computer Science Web of Science Categories. The final query is given and briefly explained in Appendix A.2.

Table 1: Selected keywords to be used in the search query

Vision	*visual, image, video, facial*
Affect	*emotion, affect, mood, feeling, facial expression*
Recognition	*detection, recognition, prediction, estimation*

2.5 Selection Process

We implemented the eligibility criteria from 2.2 to select the papers in multiple stages.

The first round of selection was done as much as possible automatically through the query itself by including some of the eligibility criteria as part of the query as described in 2.4. During the selection process each paper is marked as either "IN", "OUT" or "?", the last of which represents papers for which their eligibility is unclear in the given stage. Such papers will be reconsidered in the next stage. The final decision was made when reading the full text.

In the next stage, we screened the papers by title to remove the clear outliers that are irrelevant for this review. This includes studies using physiological signals or other different modalities than visual. Here we mainly looked at the exclusion criteria, as you cannot easily judge from the title if *all* the inclusion criteria are satisfied, but it is easier to see if *any* of the exclusion criteria are satisfied. This removed 37 papers.

Next, we screened the papers by abstract. In the interest of time, this was done in sets of 100 papers at a time to ensure enough time for reading and extracting data. Here both the exclusion and inclusion criteria were considered as much as possible. As before, if it was clear from the abstract that all inclusion criteria were satisfied, the paper is included. Otherwise, or if any of the exclusion criteria were satisfied, it is excluded. If still unclear the paper is marked as "?" to be determined in the next stage. This removed 14 papers from the first set of 100 papers. In the end, time did not allow to screen more sets, so only the remaining 86 papers of the first set are considered in the next stage.

Lastly, the paper was screened by the full text. This was combined with data extraction, if a paper was deemed eligible, then the data was also immediately extracted from the paper. Given the limited time frame, 55 records were able to be screened by full text of which 2 were excluded due to their irrelevance to the topic. In the end, we had 53 records for the review.

2.6 Search Results

The query in A.2 retrieved 549 results. Using WoS filters to exclude papers that are not in the computer science field removed 211 results. From the remaining 338 results, 53 were excluded after screening by title leaving 285 results. Then, as described in 2.4, records were screened by abstract in sets of 100 at a time, following WoS's ordering by relevance. In the interest of time only the first set was screened, of which 14 records were excluded. Of these 86 records, 55 were screened by full text, of which 2 were excluded for being out of scope. In the end we reviewed 53 records. Figure 2 shows a visualization of this filtering process.

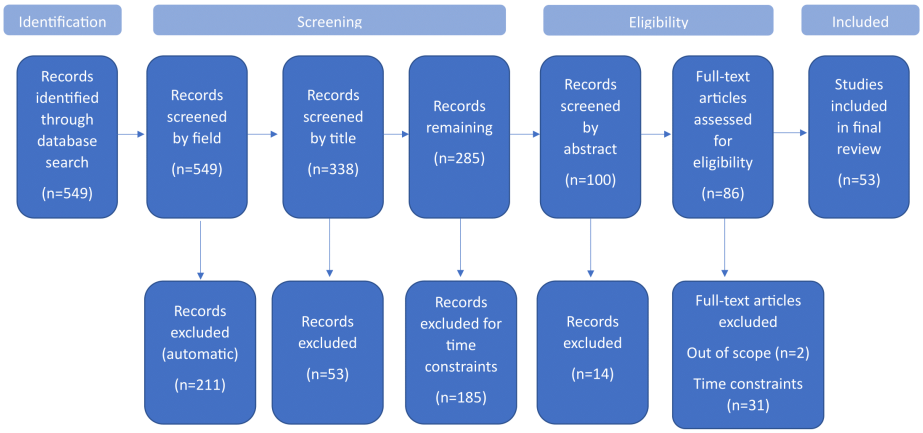


Figure 2: PRISMA diagram summarizing filtering process from the initial identification stage of 549 papers to the final 53 included papers.

2.7 Data Extraction

The data extraction process was done in parallel with screening the papers by full text, the final stage of the previously described selection process. For each paper, relevant data for each research question was extracted. This included ARS and datasets used, the motivations behind these choices, and any psychological theories provided to explain the ARS. Each paper was marked with an identifier representing different ARS, to make data aggregation at the end easier. These markers were later aggregated into the classification system used in section 3. Each paper was also labelled with the data sets used. A full overview of the collected papers and their markers can be found in Table 5 in Appendix C.

3 Results

Of the 549 papers we initially started with, we narrowed down the selection to 53 papers to include in this review. The results of this review for each sub question are summarized in this section. Several additional tables summarizing the data are included in Appendix C.

3.1 Types of affective states

What types of affective states have been targeted by prediction systems?

The types of affective states referred to here are those described by Scherer in [1], namely , attitudes, moods, affect dispositions, interpersonal stances, aesthetic emotions and utilitarian emotions. All except one of the papers in the review target utilitarian emotions, such as happiness, sadness, anger, etc. The exception to this uses facial expressions to gauge students' engagement in online classroom settings, which we describe as a mood [23].

To explore this question further, we looked at the individual affective states being targeted. As shown in Figure 3, *joy, sadness, disgust, surprise, fear,* and *anger* are the most popular states across all papers. Followed by *neutral*, used by 52% of the papers, and *contempt* being used by 24%. The rest of the states —*relief, uncertainty, squint, scream, puzzlement*, and *boredom*—are not as popular, only being used once or twice.

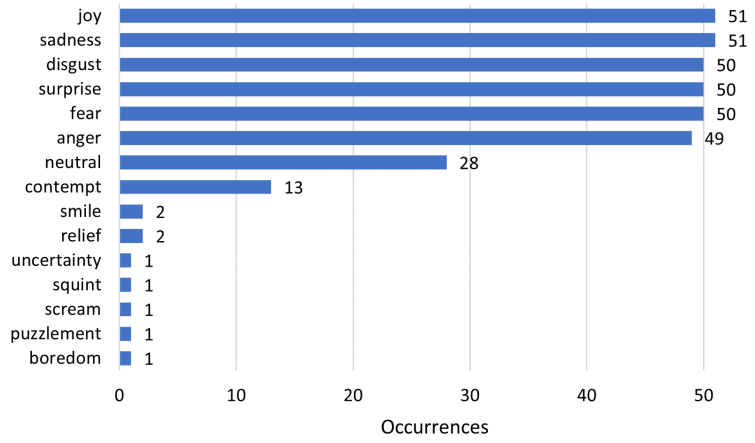


Figure 3: Occurrences of targeted affective states

3.2 Affect representation schemes

What different affect representation schemes have been used for this, and if so, what is the motivation for this particular scheme?

As stated earlier in section 1, affect representation schemes can be grouped into two categories: (1) categorical and (2) dimensional. We found 12 categorical schemes and 1 dimensional scheme. Below we describe our classification system of the ARS we encountered, and provide the reasoning provided (if any) for the use of each ARS.

Of the 12 categorical schemes, 8 were based on Ekman's Basic Emotions Theory (BET) [12]. The naming convention we chose comes from the fact that these are all based on BET. The number indicates how many states the ARS has, for those with 7 states, we use *N* or *C* to differentiate between an ARS with 7 states of which one is *neutral* or *contempt*. BET-7N-*A* indicates that this ARS is the same as BET-7N, but without *anger*. Similarly -*D* indicates the lack of *disgust*.

BET-6 : *joy, fear, sadness, surprise, disgust, anger*
Reasoning: These are considered universal human emotions across cultures [12]. Not all papers using this ARS gave explicit reasoning.

BET-6-I : BET-6 with an additional label for intensity (*low, normal, high, very high*).
Reasoning: Distinguishing between different levels of intensity allows the model to estimate in a way that is more similar to how humans recognize emotions [24].

BET-6-G/U : Two sets of BET-6: (1) The affect displayed by the subject is genuine (G), and (2) the affect displayed is unfelt (U). This gives 12 affective states in total.
Reasoning: People's facial expressions when faking an emotion can have slight differences than when the emotion being displayed matches with the emotion being felt [25].

BET-7N : *joy, fear, sadness, surprise, disgust, anger, neutral*
Reasoning: Same as BET-6, extended with a *neutral* state that can be used as a baseline.

BET-7N-A : *joy, fear, sadness, surprise, disgust, neutral*
Reasoning: Paper gave no reasoning [26].

BET-7N-D : *joy, fear, sadness, surprise, anger, neutral*
Reasoning: Paper gave no reasoning [23].

BET-7C : *joy, fear, sadness, surprise, disgust, anger, contempt*
Reasoning: After establishing the BET-6 categories, Ekman's later research findings proposed that *contempt* is also a universal emotion. Not all papers using this ARS gave explicit reasoning.

BET-8 : *joy, fear, sadness, surprise, disgust, anger, contempt, neutral*
Reasoning: Same as BET-7C, extended with a *neutral* state that can be used as a baseline.

The remaining categorical schemes deviate more from BET. Here the naming convention is similar in that the number represents the amount of affective states. The letter represents one of the unique states to distinguish between the ARS with the same number of states.

4 : *disgust, smile, surprise, sadness*
Reasoning: The paper [27] uses the CK [6] and JAFFE [7] datasets, which both use BET-6 labels. It is not explained why fear and anger are excluded or why the "happy" label is changed to "smile".

5B : *disgust, happiness, boredom, puzzlement, uncertainty*
Reasoning: This paper [28] is using a subset of the FABO [29] dataset which originally has BET-6 extended with anxiety, boredom and uncertainty. Interestingly, "puzzlement" is not one of FABO's categories. The authors give no reasoning as to why they chose this particular dataset nor why they selected this set of categories.

5R : *joy, anger, fear, sadness, relief.*
Reasoning: Unlike the rest of the categories, relief is not usually associated with Ekman's BET. This state was added to provide a balance between the positive and negative emotions, since the only other positive emotion was *joy* [30], [31].

5S : *disgust, smile, surprise, squint, scream*
Reasoning: The paper uses the *Multi-PIE* dataset [32] which uses these categories and was created as an extension to the *PIE* dataset [33]. Both of these were originally intended to be used for automatic facial recognition, not AAR. The labels are an objective description of the expression (with the exception of *disgust*). The paper gave no clear reasoning as to why this dataset was used over another.

Finally, we encountered one dimensional model. Here the naming convention is simply the name of the model.

Fontaine : *valence, arousal, power, expectancy* [34].
Reasoning: The authors in [35] chose this as one of their schemes to have the ability to represent the way in which one can express more than one emotion at simultaneously.

Table 2 shows an overview of the ARS used in the reviewed papers.

The authors of the reviewed papers don't often give clear or explicit justification for their choice of ARS. Oftentimes the choice of ARS is determined by the datasets used for training and evaluation. The choice of datasets is also not motivated by most papers. However, the dataset papers themselves do usually explain the reasoning for the used schemes.

Are systems using more than one ARS simultaneously, and if so, what is their motivation for doing so?

None of the papers use more than one ARS simultaneously to classifying the affective states. It is the case, however, that 16 of the 53 papers (30%) use multiple data sets for a larger pool of training data and during the evaluation of their systems. Their systems are then capable of identifying different affective states based on multiple ARS depending on what their training data is.

3.3 Popularity of ARS

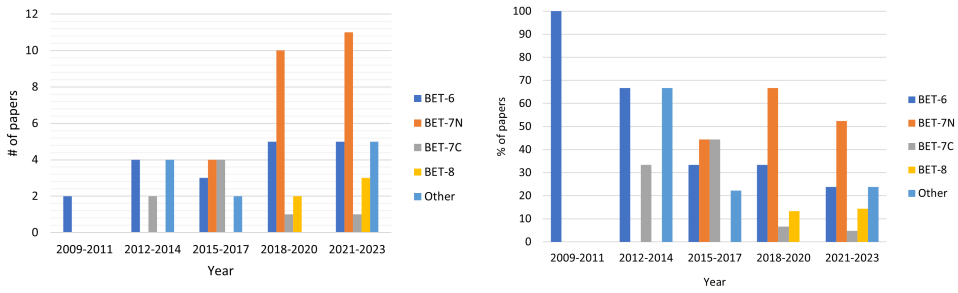
Are there differences in the popularity of schemes used for modeling different affective states?

All reviewed papers use categorical schemes, with two using an additional dimensional scheme as well. The following four schemes are the most popular categorical ARS we encountered: (1) BET-7N used in 26 papers, (2) BET-6 used in 20 papers, (3) BET-7C used in 8 papers, and (4) BET-8 used in 4 papers. BET-6 and BET-7N together clearly dominate the research area. There are 9 other schemes used throughout the reviewed papers, but these are only used once or twice. The popularity for each ARS and the papers associated are summarized in Table 2.

Has the popularity of specific schemes changed over time?

The publishing years of the reviewed papers range from 2009 to 2023, giving us a 14 year time range to analyse. In these years, BET-7N has risen in popularity while BET-6's popularity has dropped. BET-6 is the most consistently used ARS over time. BET-8 only appeared in papers published after 2018. There is more variety in ARS used after 2015 than before. This is summarized in Figure 4.

It should be noted however that the sample size of papers might not be representative for every year. The sample size of papers published in 2009-2016 is much smaller (12) than those published in 2017-2023 or later (41). This means that the data for earlier papers might not be as representative and thereby not capture the true state of the range and popularity of ARS used. This is reflected in Table 3 in Appendix C.



(a) Absolute values

(b) Relative values

Figure 4: Popularity of ARS over time shown in absolute values in 4a and in relative values in 4b. Description of each series is given in Table 2. Sample sizes in 2018-2023 is greater and thus more representative than 2009-2017. This is reflected in Table 3 in Appendix C

Table 2: ARS used in reviewed papers

Model type	ARS code	Affective states / dimensions	# papers	Associated papers
	5R	joy, anger, fear, sadness, relief	2	[30], [31]
	BET-6	joy, anger, fear, sadness, surprise, disgust	20	[31], [35]–[44] [24], [45]–[52]
	BET-7N	joy, anger, fear, sadness, surprise, disgust, neutral	26	[53]–[64] [46], [51], [65]–[76]
	BET-7C	joy, anger, fear, sadness, surprise, disgust, contempt	8	[30], [40], [53], [55], [58], [72], [77], [78]
	BET-8	joy, anger, fear, sadness, surprise, disgust, contempt, neutral	4	[62], [64], [68], [73]
Categorical	Misc	*See below*	7	
	4	disgust, smile, surprise, sadness		[27]
	5S	disgust, smile, surprise, squint, scream		[77]
	5B	disgust, happiness, boredom, puzzlement, uncertainty		[28]
	BET-6-I	BET-6 with intensity: (low, normal, high, very high)*		[24]
	BET-6-G/U	BET-6 genuine and BET-6 unfelt**		[25]
	BET-7N-A	joy, fear, sadness, surprise, disgust, neutral		[26]
	BET-7N-D	joy, fear, sadness, surprise, anger, neutral		[23]
Dimensional	Fontaine	valence, arousal, power, expectancy	2	[35], [42]

3.4 Basis in psychology

Are the majority of representation schemes used based on psychological theory?

None of the reviewed papers give any extensive psychological background or reasoning behind the ARS used. At most, the introduction briefly mentions Ekman's research and his discrete models [12]. A small minority also mention dimensional models such as Russell's valence-arousal model [13], which makes sense since only two papers actually used a dimensional model.

All reviewed papers used datasets for training and evaluation of their systems. While the reviewed paper itself might not motivate the chosen ARS or dataset, the dataset's paper does oftentimes provide more information on the chosen labels or dimensions. The most popular datasets over all reviewed papers (whose papers are available) are CK+ [6], JAFFE [7], MMI [79], FER2013 [80], and BU3DFE [81]. The CK, CK+, and MMI data sets are based on Ekman's BET and the accompanying Facial Action Coding System (FACS) used to classify a facial expression as a representation of a certain emotion. CK+ is an extension of CK to

reflect the addition of *contempt* to Ekman's BET following later research. The other two data sets, FER2013 and BU3DFE, do not provide any psychological basis for their choice of ARS. The paper introducing the FER2013 database did state that the ARS used follows another database (Toronto Face DB), but its source paper was unfortunately unavailable.

4 Discussion

The results in section 3 show clear trends in the current state of ARS usage in vision-based affect prediction. Categorical ARS, specifically BET-6 and BET-7N, dominate the research area, and there is a general lack of reasoning provided by the authors when it comes to the choice of ARS.

Our results show that all the reviewed papers use a categorical ARS for their systems, bringing us to the conclusion that this has become the norm in vision-based AAR. Since vision-based AAR usually relies on facial expressions, we believe the choice for categorical makes sense due to these being easier to categorize than to describe over dimensions, at least from a human perspective.

It has become the norm to classify affective states using BET-6 or BET-7N. Ekman proposed the 6 prototypic emotions that make up BET-6 in the 1960s, but later in the 1980s he proposed adding *contempt* to the list, making BET-7C, following further research of his. It is interesting to see that despite all of the reviewed papers being published well after this addition (the earliest being 2009), only 22% chose to include this state in their models. The majority sticks to the "old school" BET-6 or BET-7N, ignoring *contempt* completely.

One might say that even though the papers are published after 2009, the datasets used may have been released earlier thereby explaining the exclusion of *contempt*. The majority of the commonly used datasets mentioned in 3.4 were published in the 2000s or later. The exception to this is JAFFE, released in 1998. Only one of these commonly used datasets, namely CK+, includes *contempt*. This is a point of concern to us as it gives the impression that researchers are not staying up to date with developments in psychology, which should be crucial when AAR is a field that ia so closely linked.

We can only guess at the reasoning for the authors' choices due to the lack of concrete reasoning provided in these papers. This in and of itself is already a cause for concern, even if the dataset papers do provide further reasoning. We believe it is also the responsibility of the researchers using the datasets to explicitly (re)establish *why* they chose these datasets, categories, or dimensions, for their systems. Explaining why the chosen ARS is useful for their specific application or approach can be of help for future researchers building off of this work. Additionally, it shows there was critical and conscious thought behind their decisions, making their work more credible and reliable.

5 Responsible Research

As with all research, it is essential to examine the ethical implications and impact of our methods.

5.1 Reflecting on the methodology

The nature of a systematic literature review ensures reproducibility of results. As such, if one were to repeat this literature review following our methodology described in section 2, they should find the same results. However, it is important to note here that systematic literature reviews usually span months, whereas we were limited to less than 10 weeks. This limited time frame may affect the rigour and detail in comparison with longer systematic literature reviews. Additionally, as there was only one person performing the review, there is the potential for small errors in the selection and interpretation of papers. This issue could be mitigated by getting other researchers to perform the selection criteria on a randomly selected subset of papers to see if they agree with the application of the selection process. Unfortunately, given time constraints, this was not possible for this review.

Furthermore, the sample of papers retrieved for the review might not be fully representative of the state of the field of vision-based AAR. Some of the exclusion criteria, specifically limiting the results to only papers that use one of the databases listed in Appendix B, are likely to have introduced some selection bias in the results. As we did not use an exhaustive list of all the databases found in the literature, it is very likely that relevant papers were excluded. This could have affected the results in some way. There is realistically never a way to know for sure if the sample of papers is representative of the entire field, but this does not invalidate the results of this review. Our findings are still representative of at least a subset of the research area.

5.2 Ethical implications of AAR

This review is intended to aid future researchers be better informed about the current state of vision-based AAR. Understanding the decisions made within this field can help them critically improve upon or develop their own systems. However, the findings of this research can lead to development of unintentionally problematic or malicious systems.

AAR is based on predictions, and every prediction task comes with big risks of misinterpretation, which can have important consequences. When analysing facial expressions as part of an AAR system, it's important to remember that recognizing a facial expression is not the same as recognizing an emotion. Facial expression recognition is merely an estimation of the underlying emotion of the subject, and in some cases may not even match up. A subject could be hiding their true emotion, faking a facial expression. A person expressing a happy face is not necessarily a person who is happy. Additionally, classification tasks rely on generalizations. However when it comes to facial expressions and emotions, there are many cultural differences but also person-to-person differences that need to be considered. Some datasets take this into account through including subjects of various ages, ethnicities, and cultures in their data. Not all datasets do this though, so researchers should be aware of this when choosing or creating datasets for their systems.

There are many potential applications for vision-based AAR. Healthcare, educational aids, and vehicle safety [3], are generally considered to not have many negative implications. On the other hand, using AAR in areas such as safety & security can have serious negative consequences. Take for example automatic deception detection [82], [83]. This could be used in border control or police investigations [84], where misclassification can have major negative consequences for an individual. This exemplifies one of the main controversies debated when it comes to the ethical implications of AAR.

As AI continues to develop, governments are trying to mitigate the associated risks by implementing rules and regulations. The European Union (EU) published the Artificial Intelligence Act (AI Act) with guidelines for development and usage of AI systems [85]. Some sources believe that these regulations are not strict enough, especially concerning emotion recognition, claiming that the AAR systems have the potential of "undermining human rights, such as the rights to privacy, to liberty and to a fair trial" [86]. For this reason, I believe we are at a turning point where we should keep an eye on new developments in AI and critically think about every decision being made, and avoid blindly trusting these systems and their models.

6 Conclusions and Future Work

In this paper we performed a systematic literature review with the goal of exploring the current state of affect representation schemes used in vision-based automatic affect recognition systems. We looked into the types of affective states targeted, the affect representation schemes being used, their popularity, and their psychological basis. We reviewed 53 records following the PRISMA guidelines for systematic reviews [20]. The papers were selected from 549 records retrieved from WebOfScience using certain eligibility criteria to ensure the included papers were relevant to the study. These criteria included limiting our search to computer science papers written in English proposing a vision-based AAR system targeting single subjects. Studies dealing exclusively with micro-expressions or group affect recognition were excluded. Additionally, given the time limitation imposed on this research we excluded papers that were not readily accessible with our TU Delft license, used multimodal input, or did not use one of the datasets listed in Appendix B.

Our results are summarized as follows. We found that since 2009, categorical models—those that classify affective states into discrete categories—are the norm. More specifically, Ekman's Basic Emotion Theory is the basis for the most commonly used schemes, classifying facial expressions as *joy, anger, fear, sadness, surprise,* and *disgust* [12]. This is most often extended with the *neutral* state, sometimes the *contempt* state, and occasionally with both. These four are the most commonly used representations across the reviewed literature. A clear explanation of the authors' reasoning for certain ARS was difficult to find for the majority of papers, which gives the impression that researchers do not give critical thought to these decisions. This is problematic and hinders future development of AAR systems that build upon the current research. Despite the lack of motivations provided by the authors of our reviewed papers, the source papers for the used datasets gave more insights on some of the ARS, most of which are based in psychological theories.

We have several recommendations for future expansions on our research. Firstly, we recommend increasing the sample size of the set of reviewed papers. This allows for the data to be more representative of the current state of the research area. Especially increasing the number of papers from earlier years (2009-2017) to have an uniform distribution of papers over all years, as much as possible. Next, if feasible, we recommend removing the feasibility constraints. This will remove possible biases introduced when limiting the search to systems that use at least one of our set of predetermined datasets. Additionally, it allows to explore systems using multimodal input. We are interested in exploring the area of vision-*based* AAR, this does not mean vision *only*. Including multimodal inputs in this research will provide further insights on different ARS and their usage, and it would be interesting to see how this differs from purely visual input.

A Search query

A.1 Base query

(visual OR image OR video OR facial) AND
(emotion OR affect* OR mood OR feeling OR facial expression) AND
(detect* OR recogni* OR predict* OR estimat*) AND
NOT survey NOT review

A.2 Final query

Below is the final query used to retrieve the papers from WebOfScience. It is the base query extended with the data sets in order to narrow down results for feasibility within the given time constraints.

(visual OR image OR video OR facial) AND
(emotion OR affect* OR mood OR feeling OR facial expression) AND
(detect* OR recogni* OR predict* OR estimat*) AND

("CK+" OR "JAFFE" OR "BU-3DFE" OR "FER-2013" OR "EmotiW" OR "MMI" OR "eNTERFACE" OR "KDEF" OR "RaFD" OR "C-K" OR "CK" OR "BU-4DFE" OR "NVIE" OR "Affective-MIT Facial Expression" OR "DISFA" OR "LIRIS-ACCEDE" OR "FABO" OR "Kinect FaceDB")

NOT survey NOT review

The asterisk in *affect** allows to match results on *affect*, *affective*, and similar relevant words with *affect* as its root. Similarly for *detect**, *recogni**, and *estimat**, it allows for matching on both the noun (*detection, recognition, estimation, estimate(s)*) and verb (*detecting, recognizing, estimating*) versions of the word. With this base query solidified I moved on to gathering the final set of papers.

B Datasets

Below are the databases considered for limiting the search results gathered from other literature reviews on AAR [19], [21].

CK+
JAFFE
BU-3DFE
FER-2013
EmotiW
MMI
eNTERFACE
KDEF
RaFD
C-K
CK
BU-4DFE
NVIE
Affective-MIT Facial Expression
DISFA
LIRIS-ACCEDE
FABO
Kinect FaceDB

C Tables

Table 3: Popularity of ARS over time

ARS	Year 2009-2011	2012-2014	2015-2017	2018-2020	2021-2023	All-time
BET-6	2	4	3	5	5	25
BET-7N	0	0	4	10	11	25
BET-7C	0	2	4	1	1	8
BET-8	0	0	0	2	3	5
Other	0	4	2	0	5	11
Total papers	2	6	9	15	21	53

Note: Individual values per year group will not sum to "total" due to papers using multiple ARS.

Table 4: Popularity of data sets over time. Each entry shows the percentage of papers included in this study from year x that used data set y. The last row shows the total number of papers reviewed for that year so the reader can better interpret the data.

Dataset	2012	2013	2014	2016	2017	2018	2019	2020	2021	2022	2023
CK+	1	1	1	1	3	2	3	1	4	4	4
BU4DFE	1	0	0	0	0	0	0	0	1	0	0
KDEF	0	0	0	0	0	0	1	1	2	1	1
DISFA	0	0	0	0	0	1	0	0	0	0	0
FG-NET FEED	0	1	0	0	0	0	0	0	0	0	0
AFEW (eval only)	0	0	0	0	1	0	0	0	0	0	0
MUG	0	0	0	0	1	0	0	0	2	0	0
Oulu-Casia VIS	0	0	0	0	0	0	0	0	1	0	0
RAF-DB	0	0	0	0	0	0	0	0	0	1	0
BP4D-spontaneous	0	0	0	0	0	0	0	0	1	1	0
GEMEP-FERA	0	0	0	0	0	1	0	0	0	0	0
MMI	0	1	1	1	1	1	0	0	0	2	0
BU3DFE	0	0	0	1	0	0	0	0	2	1	0
JAFFE	0	0	0	1	2	0	3	1	2	1	3
bosphorus	0	0	0	0	0	0	0	0	1	1	0
RaFD	0	0	0	0	0	0	0	0	1	0	0
EmotiW-17	0	0	0	0	0	1	0	0	0	0	0
FED-RO	0	0	0	0	0	0	0	0	0	1	0
FER2013	0	0	0	0	1	0	0	1	0	0	2
KDEF-dyn	0	0	0	0	0	0	0	0	0	1	0
AffectNet	0	0	0	0	0	0	0	1	0	1	0
Total papers	2	3	2	4	6	5	3	5	10	11	4

Table 5: Summary of collected data.

Year	# of papers	Reference	Affect Representation Scheme(s) used	Data set(s) used
2009	1	[36]	BET-6	CK, JAFFE
2011	1	[39]	BET-6	BU3DFE
2012	1	[40]	BET-6, BET-7C	CK+, BU4DFE
2013	2	[41]	BET-6	CK+, MMI, FG-NET FEED
		[77]	BET-7C, 5S	BU3DFE, Multi-PIE
2014	3	[35]	BET-6, Fontaine, continuous fontaine	CK+, MMI
		[42]	BET-6, Fontaine	CK+, AVEC 2011
		[28]	5B	FABO
2015	2	[43]	BET-6	CK+, JAFFE
		[30]	BET-7C, 5R	CK+, FERA, RUFACS
2016	2	[53]	BET-7N, BET-7C	CK+, JAFFE, MMI, BU3DFE
		[57]	BET-7N	CK+, JAFFE
2017	5	[44]	BET-6	CK+, JAFFE, FER2013
		[58]	BET-7C, BET-7N	CK+, MMI, AFEW
		[59]	BET-7N	CK+, JAFFE, MUG
		[78]	BET-7C	CK+, DISFA
		[31]	BET-6, 5R	CK+, MMI, GEMEP-FERA
2018	2	[60]	BET-7N	CK+, MMI, DISFA, GEMEP-FERA
		[61]	BET-7N	CK+, EmotiW-17
2019	6	[62]	BET-7N, BET-8	CK+, JAFFE, KDEF
		[63]	BET-7N	CK+, JAFFE
		[64]	BET-7N, BET-8	CK+, JAFFE
		[54]	BET-7N	CK+, JAFFE
		[55]	BET-7N, BET-7C	CK+, JAFFE, FER2013, SFEW
		[37]	BET-6	BU4DFE, BP4D
2020	7	[56]	BET-7N	CK+, JAFFE, KDEF, FER2013, AffectNet
		[38]	BET-6	CK+, RML, AFEW5.0
		[45]	BET-6	RML, Enterface'05, AFEW 6.0
		[65]	BET-7N	BigFaceX (CK+, BAUM1, eNTERFACE)
		[71]	BET-7N	JAFFE, BU3DFE
		[46]	BET-6, BET-7N	CK+, MMI, JAFFE, SFEW2.0
		[47]	BET-6	USTC-NVIE
2021	9	[72]	BET-7N, BET-7C	CK+, MUG
		[48]	BET-6	CK+, KDEF, JAFFE, MUG
		[24]	BET-6-I	CK+, BU3DFE
		[49]	BET-6	Bosphorus, BU3DFE, BU4DFE, BP4D-spontaneous
		[73]	BET-7N, BET-8	CK+, KDEF, JAFFE, Oulu-Casia VIS, RaFD
		[74]	BET-7N	JAFFE, KDEF
		[25]	BET-6-G/U	CK+, BP4D, Oulu-Casia, SASE-FE
		[75]	BET-7N	KDEF, JAFFE
		[27]	4	CK, JAFFE
2022	6	[50]	BET-6	CK+, MMI, KDEF-dyn
		[51]	BET-6, BET-7N	CK+, Bosphorus, BU3DFE, MMI, BP4D-spontaneous
		[76]	BET-7N	CK+, KDEF, JAFFE
		[66]	BET-7N	CK+, RAF-DB, AffectNet, FED-RO
		[67]	BET-7N	CK+, KDEF
		[68]	BET-7N, BET-8	CK+, JAFFE, FER2013, KDEF, FERG
2023	6	[87]	BET-8	CK+, FER2013
		[52]	BET-6	CK+, KDEF, JAFFE
		[26]	BET-7N-A†	CK+, JAFFE, FER2013
		[69]	BET-7N	CK+, JAFFE
		[70]	BET-7N	CK+, FER2013, JAFFE
		[23]	BET-7N-D‡	CK+, RAF-DB, FER-2013, Wider Face, custom (6 emotions)

Notes:
4: disgust, smile, surprise, sad.
5B boredom, disgust, happiness, puzzlement, uncertainty.
5R anger, fear, joy , sadness, relief.
5S smile, surprise, squint, disgust, scream.
BET-6: Ekman's Basic Emotion Theory: joy, fear, disgust, surprise, anger, sadness.
BET-6-I: BET-6 with an added label of intensity.
BET-6-G/U: Two sets of BET-6, one where the affect was genuine, and the other where the affect was not genuinely felt.
BET-7N: joy, fear, disgust, surprise, anger, sadness, neutral.
BET-7C: joy, fear, disgust, surprise, anger, sadness, contempt.
BET-8: joy, fear, disgust, surprise, anger, sadness, contempt, neutral.
Fontaine: A 4-dimensional model using power, valence, activation, and expectation as its dimensions.
† Same as BET-7N but without anger: happy, sad, surprise, disgust, neutral, fear.
‡ Same as BET-7N but without disgust: happy, sad, surprise, anger, neutral, fear.

References

[1] K. R. Scherer, "What are emotions? and how can they be measured?" *Social Science Information*, vol. 44, no. 4, pp. 695–729, Dec. 2005. DOI: 10.1177/0539018405058216. [Online]. Available: https://doi.org/10.1177/0539018405058216.

[2] M. Bouhlal, K. Aarika, R. A. Abdelouahid, S. Elfilali, and E. Benlahmar, "Emotions recognition as innovative tool for improving students' performance and learning approaches," *Procedia Computer Science*, vol. 175, pp. 597–602, 2020, The 17th International Conference on Mobile Systems and Pervasive Computing (MobiSPC),The 15th International Conference on Future Networks and Communications (FNC),The 10th International Conference on Sustainable Energy Information Technology, ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2020.07.086. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050920317865.

[3] J. Fulmer, *The value of emotion recognition technology*, Sep. 2021. [Online]. Available: https://www.itbusinessedge.com/business-intelligence/value-emotion-recognition-technology/.

[4] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259–275, 2003, ISSN: 0031-3203. DOI: https://doi.org/10.1016/S0031-3203(02)00052-3. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320302000523.

[5] A. Mehrabian and J. A. Russell, *An approach to environmental psychology.* The MIT Press, 1974, pp. 266, xii, 266–xii, ISBN: 0262130904.

[6] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, 2010, pp. 94–101. DOI: 10.1109/CVPRW.2010.5543262.

[7] M. Lyons, M. Kamachi, and J. Gyoba, *The japanese female facial expression (jaffe) dataset*, en, 1998. DOI: 10.5281/ZENODO.3451524. [Online]. Available: https://zenodo.org/record/3451524.

[8] O. Langner, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, S. T. Hawk, and A. van Knippenberg, "Presentation and validation of the radboud faces database," *Cognition &amp Emotion*, vol. 24, no. 8, pp. 1377–1388, Dec. 2010. DOI: 10.1080/02699930903485076. [Online]. Available: https://doi.org/10.1080/02699930903485076.

[9] G. Castellano, B. D. Carolis, and N. Macchiarulo, "Automatic facial emotion recognition at the covid-19 pandemic time," DOI: 10.1007/s11042-022-14050-0. [Online]. Available: https://doi.org/10.1007/s11042-022-14050-0.

[10] P. Ekman, "An argument for basic emotions," *Cognition and Emotion*, vol. 6, no. 3-4, pp. 169–200, 1992. DOI: 10.1080/02699939208411068. eprint: https://doi.org/10.1080/02699939208411068. [Online]. Available: https://doi.org/10.1080/02699939208411068.

[11] P. Ekman, "Basic emotions," *Handbook of Cognition and Emotion*, pp. 45–60, 2005. DOI: 10.1002/0470013494.ch3.

[12] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion.," *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124–129, 1971. DOI: 10.1037/h0030377.

[13] J. A. Russell, "A circumplex model of affect.," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980. DOI: 10.1037/h0077714.

[14] Y.-T. Matsuda, T. Fujimura, K. Katahira, *et al.*, "The implicit processing of categorical and dimensional strategies: An fmri study of facial emotion perception," *Frontiers in human neuroscience*, vol. 7, p. 551, Sep. 2013. DOI: 10.3389/fnhum.2013.00551.

[15] D. Canedo and A. J. R. Neves, "Facial expression recognition using computer vision: A systematic review," *Applied Sciences*, vol. 9, no. 21, 2019, ISSN: 2076-3417. DOI: 10.3390/app9214678. [Online]. Available: https://www.mdpi.com/2076-3417/9/21/4678.

[16] B. C. Ko, "A brief review of facial emotion recognition based on visual information," *Sensors*, vol. 18, no. 2, 2018, ISSN: 1424-8220. DOI: 10.3390/s18020401. [Online]. Available: https://www.mdpi.com/1424-8220/18/2/401.

[17] I. Adjabi, A. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed, "Past, present, and future of face recognition: A review," *Electronics*, vol. 9, no. 8, 2020, ISSN: 2079-9292. DOI: 10.3390/electronics9081188. [Online]. Available: https://www.mdpi.com/2079-9292/9/8/1188.

[18] A. Dzedzickis, A. Kaklauskas, and V. Bucinskas, "Human emotion recognition: Review of sensors and methods," *Sensors*, vol. 20, no. 3, 2020, ISSN: 1424-8220. DOI: 10.3390/s20030592. [Online]. Available: https://www.mdpi.com/1424-8220/20/3/592.

[19] C. Vinola and K. Vimaladevi, "A survey on human emotion recognition approaches, databases and applications," *ELCVIA: electronic letters on computer vision and image analysis*, pp. 00 024–44, 2015.

[20] M. J. Page, D. Moher, P. M. Bossuyt, *et al.*, "Prisma 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews," *BMJ*, n160, Mar. 2021, ISSN: 1756-1833. DOI: 10.1136/bmj.n160.

[21] D. Canedo and A. J. R. Neves, "Facial expression recognition using computer vision: A systematic review," *Applied Sciences*, vol. 9, no. 21, 2019, ISSN: 2076-3417. DOI: 10.3390/app9214678. [Online]. Available: https://www.mdpi.com/2076-3417/9/21/4678.

[22] Jun. 2022. [Online]. Available: https : / / support . clarivate . com / ScientificandAcademicResearch / s / article / KeyWords - Plus - generation - creation - and-changes?language=en_US.

[23] S. Gupta, P. Kumar, and R. K. Tekchandani, "Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models," *MULTIMEDIA TOOLS AND APPLICATIONS*, vol. 82, pp. 11 365–11 394, 8 Mar. 2023, ISSN: 1380-7501. DOI: 10.1007/s11042-022-13558-9.

[24] O. Ekundayo and S. Viriri, "Multilabel convolution neural network for facial expression recognition and ordinal intensity estimation," *PEERJ COMPUTER SCIENCE*, vol. 7, Nov. 2021. DOI: 10.7717/peerj-cs.736.

[25] K. Kulkarni, C. A. Corneanu, S. O. Ikechukwu, *et al.*, "Automatic recognition of facial displays of unfelt emotions," *IEEE TRANSACTIONS ON AFFECTIVE COMPUTING*, vol. 12, pp. 377–390, 2 Apr. 2021, ISSN: 1949-3045. DOI: 10.1109/TAFFC.2018.2874996.

[26] N. Banskota, A. Alsadoon, P. W. C. Prasad, A. Dawoud, T. A. Rashid, and O. H. Alsadoon, "A novel enhanced convolution neural network with extreme learning machine: Facial emotional recognition in psychology practices," *MULTIMEDIA TOOLS AND APPLICATIONS*, vol. 82, pp. 6479–6503, 5 Feb. 2023, ISSN: 1380-7501. DOI: 10.1007/s11042-022-13567-8.

[27] I. M. Revina and W. R. S. Emmanuel, "Face expression recognition using ldn and dominant gradient local ternary pattern descriptors," *JOURNAL OF KING SAUD UNIVERSITY-COMPUTER AND INFORMATION SCIENCES*, vol. 33, pp. 392–398, 4 May 2021, ISSN: 1319-1578. DOI: 10.1016/j.jksuci.2018.03.015.

[28] J. Yan, W. Zheng, M. Xin, and J. Yan, "Integrating facial expression and body gesture in videos for emotion recognition," *IEICE TRANSACTIONS ON INFORMATION AND SYSTEMS*, vol. E97D, pp. 610–613, 3 Mar. 2014, ISSN: 1745-1361. DOI: 10.1587/transinf.E97.D.610.

[29] H. Gunes and M. Piccardi, "A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior," vol. 1, Jan. 2006, pp. 1148–1153. DOI: 10.1109/ICPR.2006.39.

[30] F. la Torre, W.-S. Chu, X. Xiong, F. Vicente, X. Ding, and J. Cohn, "Intraface," IEEE, 2015, ISBN: 978-1-4799-6026-2.

[31] B. Hasani and M. H. Mahoor, "Spatio-temporal facial expression recognition using convolutional neural networks and conditional random fields," IEEE, 2017, pp. 790–795, ISBN: 978-1-5090-4023-0. DOI: 10.1109/FG.2017.99.

[32] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010, Best of Automatic Face and Gesture Recognition 2008, ISSN: 0262-8856. DOI: https://doi.org/10.1016/j.imavis.2009.08.002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885609001711.

[33] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression (pie) database," in *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, 2002, pp. 53–58. DOI: 10.1109/AFGR.2002.1004130.

[34] J. R. Fontaine, K. R. Scherer, E. B. Roesch, and P. C. Ellsworth, "The world of emotions is not two-dimensional," *Psychological science*, vol. 18, no. 12, pp. 1050–1057, 2007.

[35] A. C. Cruz, B. Bhanu, and N. S. Thakoor, "Vision and attention theory based sampling for continuous facial emotion recognition," *IEEE TRANSACTIONS ON AFFECTIVE COMPUTING*, vol. 5, pp. 418–431, 4 Oct. 2014, ISSN: 1949-3045. DOI: 10.1109/TAFFC.2014.2316151.

[36] L. Zhang and D. Tjondronegoro, "Selecting, optimizing and fusing 'salient' gabor features for facial expression recognition," C. S. Leung, M. Lee, and J. H. Chan, Eds., vol. 5863, SPRINGER-VERLAG BERLIN, 2009, pp. 724–732, ISBN: 978-3-642-10676-7.

[37] Q. Zhen, D. Huang, H. Drira, B. B. Amor, Y. Wang, and M. Daoudi, "Magnifying subtle facial motions for effective 4d expression recognition," *IEEE TRANSACTIONS ON AFFECTIVE COMPUTING*, vol. 10, pp. 524–536, 4 Oct. 2019, ISSN: 1949-3045. DOI: 10.1109/TAFFC.2017.2747553.

[38] X. Pan, "Fusing HOG and convolutional neural network spatial–temporal features for video-based facial expression recognition," *IET Image Processing*, vol. 14, no. 1, pp. 176–182, Jan. 2020. DOI: 10.1049/iet-ipr.2019.0293. [Online]. Available: https://doi.org/10.1049/iet-ipr.2019.0293.

[39] S. Berretti, B. B. Amor, M. Daoudi, and A. del Bimbo, "3d facial expression recognition using sift descriptors of automatically detected keypoints," *VISUAL COMPUTER*, vol. 27, pp. 1021–1036, 11, SI Nov. 2011, ISSN: 0178-2789. DOI: 10.1007/s00371-011-0611-x.

[40] L. A. Jeni, A. Lorincz, T. Nagy, *et al.*, "3d shape estimation in video sequences provides high precision evaluation of facial expressions," *IMAGE AND VISION COMPUTING*, vol. 30, pp. 785–795, 10 Oct. 2012, ISSN: 0262-8856. DOI: 10.1016/j.imavis.2012.02.003.

[41] R. A. Khan, A. Meyer, H. Konik, and S. Bouakaz, "Framework for reliable, real-time facial expression recognition for low resolution images," *PATTERN RECOGNITION LETTERS*, vol. 34, pp. 1159–1168, 10 Jul. 2013, ISSN: 0167-8655. DOI: 10.1016/j.patrec.2013.03.022.

[42] S. Shojaeilangari, W.-Y. Yau, and E.-K. Teoh, "A novel phase congruency based descriptor for dynamic facial expression analysis," *PATTERN RECOGNITION LETTERS*, vol. 49, pp. 55–61, Nov. 2014, ISSN: 0167-8655. DOI: 10.1016/j.patrec.2014.06.009.

[43] S. L. Happy and A. Routray, "Automatic facial expression recognition using features of salient facial patches," *IEEE TRANSACTIONS ON AFFECTIVE COMPUTING*, vol. 6, pp. 1–12, 1 Jan. 2015, ISSN: 1949-3045. DOI: 10.1109/TAFFC.2014.2386334.

[44] S. Munasinghe, C. Fookes, and S. Sridharan, "Deep features-based expression-invariant tied factor analysis for emotion recognition," IEEE, 2017, pp. 546–554, ISBN: 978-1-5386-1124-1.

[45] N. Hajarolasvadi and H. Demirel, "Iet image processing deep facial emotion recognition in video using eigenframes," *IET Image Process*, vol. 14, pp. 3536–3546, 14 2020, ISSN: 1751-9659. DOI: 10.1049/iet-ipr.2019.1566. [Online]. Available: www.ietdl.org.

[46] I. Gogic, M. Manhart, I. S. Pandzic, and J. Ahlberg, "Fast facial expression recognition using local binary features and shallow neural networks," *VISUAL COMPUTER*, vol. 36, pp. 97–112, 1 Jan. 2020, ISSN: 0178-2789. DOI: 10.1007/s00371-018-1585-8.

[47] G. Pons, A. E. Ali, and P. Cesar, "Et-cyclegan: Generating thermal images from images in the visible spectrum for facial emotion recognition," International Conference on Multimodal Interaction (ICMI), ELECTR NETWORK, OCT 25-29, 2020, ASSOC COMPUTING MACHINERY, 2020, pp. 87–91, ISBN: 978-1-4503-8002-7. DOI: 10.1145/3395035.3425258.

[48] F. Ayeche and A. Alti, "Facial expressions recognition based on delaunay triangulation of landmark and machine learning," *TRAITEMENT DU SIGNAL*, vol. 38, pp. 1575–1586, 6 Dec. 2021, ISSN: 0765-0019. DOI: 10.18280/ts.380602.

[49] M. Behzad, N. Vo, X. Li, and G. Zhao, "Towards reading beyond faces for sparsity-aware 3d/4d affect recognition," *NEUROCOMPUTING*, vol. 458, pp. 297–307, Oct. 2021, ISSN: 0925-2312. DOI: 10.1016/j.neucom.2021.06.023.

[50] S. Pitchaiyan and N. Savarimuthu, "Deep stacked autoencoder-based automatic emotion recognition using an efficient hybrid local texture descriptor," *JOURNAL OF INFORMATION TECHNOLOGY RESEARCH*, vol. 15, 1 2022, ISSN: 1938-7857. DOI: 10.4018/JITR.2022010103.

[51] E. Owusu, J. K. Appati, and P. Okae, "Robust facial expression recognition system in higher poses," *VISUAL COMPUTING FOR INDUSTRY BIOMEDICINE AND ART*, vol. 5, 1 May 2022. DOI: 10.1186/s42492-022-00109-0.

[52] N. H. N. Kumar, A. S. Kumar, G. M. S. Prasad, and M. A. Shah, "Automatic facial expression recognition combining texture and shape features from prominent facial regions," *IET IMAGE PROCESSING*, vol. 17, pp. 1111–1125, 4 Mar. 2023, ISSN: 1751-9659. DOI: 10.1049/ipr2.12700.

[53] L. Zhang, K. Mistry, S. C. Neoh, and C. P. Lim, "Intelligent facial emotion recognition using moth-firefly optimization," *KNOWLEDGE-BASED SYSTEMS*, vol. 111, pp. 248–267, Nov. 2016, ISSN: 0950-7051. DOI: 10.1016/j.knosys.2016.08.018.

[54] R. I. Bendjillali, M. Beladgham, K. Merit, and A. Taleb-Ahmed, "Improved facial expression recognition based on dwt feature for deep cnn," *ELECTRONICS*, vol. 8, 3 Mar. 2019, ISSN: 2079-9292. DOI: 10.3390/electronics8030324.

[55] Y. Wang, Y. Li, Y. Song, and X. Rong, "Facial expression recognition based on auxiliary models," *ALGORITHMS*, vol. 12, 11 Nov. 2019. DOI: 10.3390/a12110227.

[56] Z. Fei, E. Yang, D. D.-U. Li, *et al.*, "Deep convolution network based emotion analysis towards mental health care," *NEUROCOMPUTING*, vol. 388, pp. 212–227, May 2020, ISSN: 0925-2312. DOI: 10.1016/j.neucom.2020.01.034.

[57] V. Mayya, R. M. Pai, and M. M. M. Pai, "Automatic facial expression recognition using dcnn," J. Mathew, D. DasKrishna, and J. Jose, Eds., vol. 93, ELSEVIER SCIENCE BV, 2016, pp. 453–461. DOI: 10.1016/j.procs.2016.07.233.

[58] X. Fan and T. Tjahjadi, "A dynamic framework based on local zernike moment and motion history image for facial expression recognition," *PATTERN RECOGNITION*, vol. 64, pp. 399–406, Apr. 2017, ISSN: 0031-3203. DOI: 10.1016/j.patcog.2016.12.002.

[59] R. Jiang, A. T. S. Ho, I. Cheheb, N. Al-Maadeed, S. Al-Maadeed, and A. Bouridane, "Emotion recognition from scrambled facial images via many graph embedding," *PATTERN RECOGNITION*, vol. 67, pp. 245–251, Jul. 2017, ISSN: 0031-3203. DOI: 10.1016/j.patcog.2017.02.003.

[60] M. Verma, J. K. Bhui, S. K. Vipparthi, and G. Singh, "Expertnet: Exigent features preservative network for facial expression recognition," ASSOC COMPUTING MACHINERY, 2018, ISBN: 978-1-4503-6615-1. DOI: 10.1145/3293353.3293374.

[61] F.-J. Chang, A. T. Tran, T. Hassner, I. Masi, R. Nevatia, and G. Medioni, "Expnet: Landmark-free, deep, 3d facial expressions," IEEE, 2018, pp. 122–129, ISBN: 978-1-5386-2335-0. DOI: 10.1109/FG.2018.00027.

[62] R. Melaugh, N. Siddique, S. Coleman, and P. Yogarajah, "Facial expression recognition on partial facial sections," S. Loncaric, R. Bregovic, M. Carli, and M. Subasic, Eds., IEEE, 2019, pp. 193–197, ISBN: 978-1-7281-3140-5.

[63] J.-H. Kim, B.-G. Kim, P. P. Roy, and D.-M. Jeong, "Efficient facial expression recognition algorithm based on hierarchical deep neural network structure," *IEEE ACCESS*, vol. 7, pp. 41 273–41 285, 2019, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2907327.

[64] Y. Miao, H. Dong, J. M. A. Jaam, and A. E. Saddik, "A deep learning system for recognizing facial expression in real-time," *ACM TRANSACTIONS ON MULTIMEDIA COMPUTING COMMUNICATIONS AND APPLICATIONS*, vol. 15, 2 Jun. 2019, ISSN: 1551-6857. DOI: 10.1145/3311747.

[65] J. R. H. Lee and A. Wong, "Timeconvnets: A deep time windowed convolution neural network design for real-time video facial expression recognition," IEEE COMPUTER SOC, 2020, pp. 9–16, ISBN: 978-1-7281-9891-0. DOI: 10.1109/CRV50864.2020.00010.

[66] K. Babu, C. Kumar, and C. Kannaiyaraju, "Face recognition system using deep belief network and particle swarm optimization," *INTELLIGENT AUTOMATION AND SOFT COMPUTING*, vol. 33, pp. 317–329, 1 2022, ISSN: 1079-8587. DOI: 10.32604/iasc.2022.023756.

[67] H. A. Shehu, W. N. Browne, and H. Eisenbarth, "An anti-attack method for emotion categorization from images," *APPLIED SOFT COMPUTING*, vol. 128, Oct. 2022, ISSN: 1568-4946. DOI: 10.1016/j.asoc.2022.109456.

[68] T. Dar, A. Javed, S. Bourouis, H. S. Hussein, and H. Alshazly, "Efficient-swishnet based system for facial emotion recognition," *IEEE ACCESS*, vol. 10, pp. 71 311–71 328, 2022, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2022.3188730.

[69] X. Guo, S. Lu, S. Wang, Z. Lu, and Y. Zhang, "Dlsanet: Facial expression recognition with double-code lbp-layer spatial-attention network," *IET IMAGE PROCESSING*, Jun. 2023, ISSN: 1751-9659. DOI: 10.1049/ipr2.12817.

[70] T. Podder, D. Bhattacharya, P. Majumder, and V. E. Balas, "A feature boosted deep learning method for automatic facial expression recognition," *PEERJ COMPUTER SCIENCE*, vol. 9, Jan. 2023. DOI: 10.7717/peerj-cs.1216.

[71] Y. Zhao and D. Chen, "A facial expression recognition method using improved capsule network model," *SCIENTIFIC PROGRAMMING*, vol. 2020, Oct. 2020, ISSN: 1058-9244. DOI: 10.1155/2020/8845176.

[72] J. M. Lopez-Gil and N. Garay-Vitoria, "Photogram classification-based emotion recognition," *IEEE ACCESS*, vol. 9, pp. 136 974–136 984, 2021, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3117253.

[73] M. Kas, Y. E. Merabet, Y. Ruichek, and R. Messoussi, "New framework for person-independent facial expression recognition combining textural and shape analysis through new feature extraction approach," *INFORMATION SCIENCES*, vol. 549, pp. 200–220, Mar. 2021, ISSN: 0020-0255. DOI: 10.1016/j.ins.2020.10.065.

[74] V. G. V. Mahesh, C. Chen, A. N. J. R. Vijayarajan, Raj, and P. T. Krishnan, "Shape and texture aware facial expression recognition using spatial pyramid zernike moments and law's textures feature set," *IEEE ACCESS*, vol. 9, pp. 52 509–52 522, 2021, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3069881.

[75] M. A. H. Akhand, S. Roy, N. Siddique, M. A. S. Kamal, and T. Shimamura, "Facial emotion recognition using transfer learning in the deep cnn," *ELECTRONICS*, vol. 10, 9 May 2021. DOI: 10.3390/electronics10091036.

[76] R. Angeline and A. A. Nithya, "Multimodal human facial emotion recognition using densenet-161 and image feature stabilization algorithm," *TRAITEMENT DU SIGNAL*, vol. 39, pp. 2165–2172, 6 Dec. 2022, ISSN: 0765-0019. DOI: 10.18280/ts.390630.

[77] X. Huang, G. Zhao, and M. Pietikainen, "Emotion recognition from facial images with arbitrary views," T. Burghardt, D. Damen, W. MayolCuevas, and M. Mirmehdi, Eds., B M V A PRESS, 2013. DOI: 10.5244/C.27.76.

[78] R. Walecki, O. Rudovic, V. Pavlovic, and M. Pantic, "Variable-state latent conditional random field models for facial expression analysis," *IMAGE AND VISION COMPUTING*, vol. 58, pp. 25–37, Feb. 2017, ISSN: 0262-8856. DOI: 10.1016/j.imavis.2016.04.009.

[79] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *2005 IEEE International Conference on Multimedia and Expo*, 2005, 5 pp.-. DOI: 10.1109/ICME.2005.1521424.

[80] I. J. Goodfellow, D. Erhan, P. L. Carrier, *et al.*, "Challenges in representation learning: A report on three machine learning contests," in *Neural Information Processing: 20th International Conference, ICONIP 2013, Daegu, Korea, November 3-7, 2013. Proceedings, Part III 20*, Springer, 2013, pp. 117–124.

[81] L. Yin, X. Wei, Y. Sun, J. Wang, and M. Rosato, "A 3d facial expression database for facial behavior research," in *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, 2006, pp. 211–216. DOI: 10.1109/FGR.2006.6.

[82] H. Bouma, G. Burghouts, R. den Hollander, *et al.*, "Measuring cues for stand-off deception detection based on full-body nonverbal features in body-worn cameras," in *Optics and Photonics for Counterterrorism, Crime Fighting, and Defence XII*, D. Burgess, G. Owen, H. Bouma, F. Carlysle-Davies, R. J. Stokes, and Y. Yitzhaky, Eds., International Society for Optics and Photonics, vol. 9995, SPIE, 2016, 99950N. DOI: 10.1117/12.2241183. [Online]. Available: https://doi.org/10.1117/12.2241183.

[83] D. Avola, L. Cinque, G. L. Foresti, and D. Pannone, "Automatic deception detection in rgb videos using facial action units," in *Proceedings of the 13th International Conference on Distributed Smart Cameras*, ser. ICDSC 2019, Trento, Italy: Association for Computing Machinery, 2019, ISBN: 9781450371896. DOI: 10.1145/3349801.3349806. [Online]. Available: https://doi.org/10.1145/3349801.3349806.

[84] T. Macaulay, *British police to trial facial recognition system that detects your mood*, Aug. 2020. [Online]. Available: https://thenextweb.com/news/british-police-to-trial-facial-recognition-system-that-detects-your-mood.

[85] *The artificial intelligence act*, Apr. 2023. [Online]. Available: https://artificialintelligenceact.eu/.

[86] A. Now, E. D. R. (EDRi), B. of Freedom, ARTICLE19, and IT-Pol, *Prohibit emotion recognition in the artificial intelligence act*, Nov. 2021.

[87] T. Shahzad, K. Iqbal, M. A. Khan, Imran, and N. Iqbal, "Role of zoning in facial expression using deep learning," *IEEE ACCESS*, vol. 11, pp. 16 493–16 508, 2023, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2023.3243850.