Affective Computing Databases: In-Depth Analysis of Systematic Reviews and Surveys

Pedro J. Vaz, João M. F. Rodrigues, and Pedro J. S. Cardoso

***Abstract*—The field of affective computing (AffC) is a hot research topic, where keeping track of the latest state-of-the-art can be cumbersome. Probably, due to this, a huge increase in publications of systematic reviews or surveys (SRoS) is appearing in different journals, covering various aspects such as databases, methods, and overall perspectives. Nevertheless, this increase does not mean more and better information, or at least a clarification of information. The present study analyses 10 SRoS, all published within the last 4 years, focusing only on covering AffC databases, with emphasis on collections where emotion or sentiment can be extracted from the body. It was observed that, depending on the SRoS, different information was presented, sometimes with missing or discrepant data, due to lack of information or by the way it was interpreted. As a result, from those 10 SRoS, a total of 111 different databases were analyzed, which were segmented into three groups (tiers, i.e., citation-based categorization) by their relative importance of appearance in the SRoS. In addition, it is proposed a taxonomy with a minimum set of characterizing information that researchers should address when publishing or reviewing databases.**

***Index Terms*—Affective computing, affective computing databases, emotion dataset, sentiment dataset.**

I. Introduction

Affective computing (AffC) is a multidisciplinary field that combines computer science, psychology, and artificial intelligence, to develop systems and technologies capable of recognizing, interpreting, and responding to human emotions and affective states [1], [2]. This research field has a wide range of applications, including human-computer interaction (HCI), healthcare, education, entertainment, and marketing. For example, AffC can be used to monitor and analyze patient emotional states in healthcare, to develop intelligent systems that adapt to analyze and respond to customer emotional responses to products and advertisements in marketing, to develop emotionally intelligent games and virtual reality systems in entertainment, to develop systems that recognize and respond to suspicious behavior in security systems etc. In summary, whenever HCI is present, AffC can be used to develop systems that can recognize and respond to the users' emotional states [3], [4], [5], [6], [7], [8].

Depending on the type and combination of data inputs used, affective computing typically uses one or a combination of the following modalities: visual information, sound/speech information, physiological signals, and textual information [9], [10], [11], [12]. Examples of visual information include facial expression, body gesture, and eye movement, while examples of speech information include speech prosody and intonation. Physiological signals include electroencephalogram (EEG), electrooculogram (EOG), electromyography (EMG), galvanic skin response (GSR), respiration rate (RESP), electrocardiogram (ECG), and heart rate (HR) [13]. Examples of textual information include social media posts, customer reviews, or forum discussions, which can be analyzed by natural language processing (NLP) techniques.

Unimodal affective computing relies on using a single source of information, such as visual or sound/speech, while multimodal involves combining information from multiple sources, such as visual and sound/speech or visual and text information. Since emotions are intricate and manifest through diverse channels, combining information from multiple modalities might yield better model performance, and is generally the preferred option by researchers. The visual modality can be further segmented by using single or sequence frames (picture *vs* video) and focusing on the face, upper body, or full body [3], [14].

Affective computing involves two distinct tasks: emotion and sentiment analysis [15], [16], [17]. To be explicit, emotion and sentiment detection deals with locating and recognizing, for example, detecting a face displaying an emotion inside a picture, whereas emotion and sentiment classification concentrates on classifying full images into predetermined classes. To simplify the writing, both terms (classification and detection) will be addressed as emotion analysis and sentiment analysis in the text, except for situations where the distinction is required for technical clarification.

Different models of human emotion have been proposed by psychologists that fall within two main categories: discrete (or categorical) and dimensional models [3], [16], [17]. As the name suggests, discrete models categorize emotions into a set of discrete categories, while dimensional models categorize emotions into a set of continuous dimensions. For example, one of the most accepted discrete models is Ekman's model [18], where emotions are categorized into a set of six basic emotions,

Received 2 May 2024; revised 16 October 2024; accepted 20 November 2024. Date of publication 27 November 2024; date of current version 27 May 2025. This work was supported by Project Sustainable Horizons, European Universities Designing the Horizons of Sustainability (SHEs) under Grant 101071300 and in part by the by NOVA LINCS ref. UIDB/04516/2020 (https://doi.org/10.54499/UIDB/04516/2020) and ref. UIDP/04516/2020 (https://doi.org/10.54499/UIDP/04516/2020) with the financial support of FCT.IP. Recommended for acceptance by J. Shukla. *(Corresponding author: Pedro J. Vaz.)*

The authors are with the NOVA LINCS & ISE, Universidade do Algarve, 8005-139 Faro, Portugal (e-mail: pjmartins@ualg.pt; jrodrig@ualg.pt; pcardoso@ualg.pt).

Digital Object Identifier 10.1109/TAFFC.2024.3507289

namely, happiness, sadness, anger, fear, disgust, and surprise. Some of Ekman's model variants also include neutral, and contempt [19], [20], [21], [22]. Another well-known discrete model is Plutchik's model [23], where emotions are categorized into eight primary emotions, specifically joy, trust, fear, surprise, sadness, disgust, anger, and anticipation, which can then be combined to form secondary emotions (e.g., admiration, terror, amazement, grief, or ecstasy). Dimensional models categorize emotions into a set of continuous dimensions, such as valence and arousal [24], [25]. In contrast, sentiment analysis involves classifying a person's opinion or behavior, which is categorized as e.g., positive, negative, or neutral [26], [27], [28], [29], [30], [31], [32]. Emotion and sentiment are crucial for fostering engagement, as they influence cognitive and behavioral responses, which are essential for maintaining meaningful and dynamic human-machine interactions [33].

In a nutshell, this study focuses on analyzing systematic reviews or surveys (SRoS) on AffC that partially or fully cover affective computing databases. The reason for this is that when analyzing recent SRoS, depending on the SRoS, different information is presented, sometimes with missing or discrepant data, either by lack of information or by the way it was interpreted. These discrepancies make difficult the comparison of databases, and the selection of the most appropriate database for a specific research objective. For instance, some SRoS do not refer the number of samples, which is a key factor in machine learning, as the models need a considerable amount of data to be successfully trained. Also, some SRoS do not mention, systematically, the number and type of emotions that are present in the database, which is another key factor in selecting the database for a specific research objective. This lack of information can lead to an initial selection of databases that are not suitable for the research objective, or to the exclusion of databases that could be useful for the research objective, possibly delaying the research progress.

In this context, this paper analyzes and discusses 10 selected SRoS papers published between 2020 and February 2024 that focus on affective computing databases. The main contributions of this paper are: i) a classification by tiers (based on citation categorization) of the databases presented in the SRoS (111 databases identified); ii) a detailed analysis and discussion of the most prevalent incorrect or absent information that appears in AffC SRoS about databases; and iii) a proposal of a taxonomy with a minimum set of characterizing information that researchers should address when publishing or reviewing databases for AffC.

The remainder of this paper is organized as follows. Section II presents the review methodology and related works. Section III describes the databases selected in the review, and Section IV introduces the proposed AffC database taxonomy. Finally, Section V presents the conclusions and future work.

II. Review Methodology & Related Works

To the best knowledge of the authors, this is the first time that an in-depth analysis of SRoS papers on AffC databases is presented. Although, Wang et al. [16] presented a section with an overview of the (at the time) latest SRoS on AffC, it was limited to a single section, had a general macro overview, and did not go into detail comparing the information presented on each SRoS against each other, since the focus of the paper was on affective computing in general, and not specifically on analyzing other SRoS. The same can be said about the other SRoS analyzed in this work, as they all present a general overview of the field without providing detailed comparisons of the information presented in each SRoS. The lack of a comparative discussion between SRoS papers makes it difficult for readers to see the discrepancies in database presentation across different SRoS, a gap that our survey addresses. As such, previous studies provide an overview of AffC databases but do not delve into the potential inconsistencies or absent information across different SRoS.

The remainder of this section presents the papers' selection criteria and a summary of the selected SRoS papers.

A. Selection Criteria

The papers studied and analyzed were selected based on the following criteria: i) the paper must be a systematic review or survey paper, ii) the paper must be published in a journal, iii) the paper must be written in English, iv) the paper must be available online, v) the paper must contain systematic information on affective computing databases, and vi) must have been published between 2020 and February 2024.

The search was conducted using academic journal databases (e.g., IEEE Transactions on Affective Computing), including IEEE Xplore, ACM Digital Library, Science Direct, Scopus, Google Scholar, and ResearchGate. The following keywords were used: "affective computing", "emotion recognition", "sentiment analysis", "affective computing databases", and "affective computing surveys". The search was performed at the end of January 2024, and a total of 20 SRoS AffC papers were identified. After applying the selection criteria, 10 SRoS papers were selected for analysis [3], [11], [12], [16], [17], [34], [35], [36], [37], [38].

B. Selected Review Papers Summary

Papers ([3], [11], [12], [16], [17], [34], [35], [36], [37], [38]) explore various aspects of AffC, including emotion recognition, sentiment analysis, and the use of different data modalities. The authors aim to provide a comprehensive review of the field, categorizing it into two broad classes: unimodal and multimodal emotion and sentiment analysis. They also discuss emotion models, databases, recent advances, and the implications of different emotion and sentiment analysis methods. As already mentioned, this study is different from the previous ones, as it is an in-depth analysis of systematic reviews and surveys, focusing on AffC databases, presenting and comparing the databases referenced in those SRoS. Next, a brief summary of each of the 10 SRoS papers that were analyzed is presented.

Li and Deng (2020) [37] provide a comprehensive overview of facial expression recognition (FER) systems, encompassing various aspects such as databases, algorithms, challenges, and applications. The paper discusses the transition of FER from lab-controlled to in-the-wild settings, the adoption of deep learning techniques to address challenges (e.g., overfitting due to insufficient training data), and expression-unrelated variations (e.g., illumination, head pose, and identity bias). The review also explores the evolution of FER algorithms and databases, including a review of advances (until 2019) in deep learning for both static and dynamic FER tasks. The survey aims to provide a systematic framework and essential skills for FER, and discusses the standard pipeline of deep FER systems, state-of-the-art neural networks and training strategies, performance evaluations, related issues, challenges, and future directions in the field.

Shoumy et al. (2020) [17] survey delves into multimodal Big Data affective analytics, particularly addressing the integration of physiological information with other modalities for sentiment and emotion recognition, such as text, audio, and visual information. The authors claim that the internet offers a wealth of data crucial for sentiment analysis (e.g., customer product reviews, social media posts, forum discussions, or blogs). They also claim that a significant portion of this data exists in unstructured and disorganized formats, referred to as Big Data. In that context, the survey provides information on several Big Data analysis frameworks that help in processing data to produce useful information. In summary, the work aims to provide information on the current state-of-the-art techniques, frameworks, fusion techniques, available databases, research works, applications and finalizing by discussing challenges, and future directions in the field of multimodal Big Data affective analytics.

Noroozi et al. (2021) [3] survey focuses on the problem of automatic emotion recognition from body gestures. The significance of nonverbal communication, often referred to as body language, and its role in conveying emotions and thoughts during conversations is discussed, including highlighting the historical and foundational aspects of body language research, as well as its relevance to human interaction. In that context, the paper provides a review of advancements in the automatic recognition of emotions through body gestures, covering various subject aspects (e.g., cultural and gender dependency), and available databases for training recognition systems. The paper also highlights challenges in the field, such as the scarcity of labeled data and the lack of agreement on clearly defined output spaces and representations.

Arya et al. (2021) [34] present a survey on multidisciplinary domains contributing to affective computing, such as sociology, psychology, computer science, physiology, mathematics, and linguistics. The aim is to explore the cumulative impact of these interrelated domains on AffC, discussing the theories, concepts, models, and implications of these fields. Additionally, the survey presents existing databases and discusses various applications where AffC has a significant impact. Furthermore, affect detection from various behavioral and physiological signals is discussed, emphasizing the challenges and advantages of each type of signal, and outlining the process of acquiring, processing, and analyzing physiological signals for AffC applications.

Baffour et al. (2022) [35] survey discusses the current state, methods, databases, and future trends in FER. The importance of emotions in human communication is highlighted, with a particular emphasis on facial expressions as a crucial aspect of non-verbal communication. It outlines various applications where FER can be utilized, including mental health detection, human-robot interaction, and market research. Furthermore, it mentions the role of deep learning algorithms in FER and reviews research works, methods, performances, efficiencies, setbacks, and open issues in FER. The survey also addresses the impact of computational power and the availability of large facial emotion databases on the pace of progress in FER research.

Jampour and Javidi (2022) [36] survey multi-view facial expression recognition (MFER), and address the challenges associated with recognizing facial expressions in varying head poses. It aims to provide a comprehensive overview of advances in MFER, covering both traditional and deep learning approaches. The paper discusses challenges such as ethnicity, head pose variation, gender, age, skin tone, and lighting conditions that complicate facial expression recognition. Additionally, it emphasizes the lack of comprehensive resources and standard validation protocols in the field of MFER. It provides a review of publicly available databases for MFER, including lab collections and databases gathered from real-world scenarios, and introduces popular validation protocols on each database to standardize comparisons in future research.

Siddiqui et al. (2022) [11] surveyed databases for multimodal emotion recognition (MER), although also providing information on existing unimodal databases. The paper discusses the significance of multimodal systems over unimodal ones, highlighting their more natural, expressive, efficient, unambiguous, and wider application domain. The use of different modalities is explored, such as facial expressions, speech, physiological signals, and body movements, in combination, to develop more accurate emotion recognition systems. The paper also emphasizes the importance of high-quality and large databases in training and verifying MER systems, particularly in the context of machine learning and deep learning algorithms. Considering this, the authors identify, describe, and classify multimodal databases and modalities used in emotion detection. Additionally, the paper introduces a new visible and infrared image database called VIRI DB for emotion recognition.

Yadav et al. (2022) [38] provide a comprehensive survey on speech and vision emotion recognition systems, machine learning architectures and algorithms, and their applications in speech and vision systems. The survey explores the state-of-the-art and discusses how machine learning can be used for learning and training with large sets of sensor data, cloud computing, and mobile and embedded technology. The challenges and success rates of using machine learning in platforms with limited resources, such as automobiles, robots, and mobile phones are highlighted. Furthermore, the demand for strong, robust, and highly intelligent systems with lower latency and high fidelity, particularly in hardware devices with resource constraints is addressed. The paper hints at emerging applications of speech and vision systems, such as evaluation, transportation, and medical prescriptions.

Wang et al. (2022) [16] presented a survey on AffC, which encompasses the recognition and analysis of human emotions and sentiments. The paper explores various aspects of affective computing, including emotion recognition, sentiment analysis, and the use of different data modalities such as text, audio, visual, and physiological signals. The authors aim to provide a comprehensive review of the field, categorizing it into two broad classes: unimodal and multimodal. They also discuss emotion models, databases, recent advances, and the implications of different emotion and sentiment analysis methods. The paper's major contributions include categorizing and taxonomizing affective computing research, providing a systematic review of numerous papers, and discussing benchmark databases and real-life applications in this domain.

Ahmed et al. (2023) [12] focus mainly on emotion recognition, in the context of machine learning and deep learning techniques. The survey provides a systematic literature review of emotion acquisition methods, emotion classification algorithms, and data fusion techniques used in emotion recognition. It also discusses fine-grained emotions, available databases, and various tools and technologies for data acquisition. Additionally, it emphasizes the efficiency of machine learning and deep learning algorithms in addressing the preprocessing and feature extraction processes in multimodal emotion recognition. The paper highlights the gaps in current research and suggests the potential use of emerging technologies like artificial intelligence, federated learning, and transfer learning for evaluating human emotion states.

In general, the selected SRoS papers provide a comprehensive overview of the field of affective computing, discussing various aspects such as emotion recognition, sentiment analysis, and the use of different data modalities. They also discuss emotion models, databases, recent advances, and the implications of different emotion and sentiment analysis methods. However, none of the papers provide a detailed analysis of the databases referenced in the SRoS papers, which is the focus of this work. In the next section, the databases presented in the 10 SRoS are characterized and segmented into tiers.

III. Databases

All datasets mentioned in the 10 SRoS ([3], [11], [12], [16], [17], [34], [35], [36], [37], [38]) are characterized and segmented into three tiers, based on the number of times they were cited in those SRoS. That is, the inclusion of the database in our study is based on the criteria that it was referenced in at least one of the selected SRoS. A total of 111 different databases were found, and in this context, the databases were segmented into three tiers: (a) Tier 1 - databases that were referenced in only one SRoS (see Table IV; to improve the paper readability, big tables were placed at the end of the paper), (b) Tier 2 - databases that were referenced in 2 or 3 SRoS (see Table V), and (c) Tier 3 - databases that were referenced at least in 4 of the selected SRoS papers (see Table VI). In this view, Tier 1 is composed of 64 databases, Tier 2 of 38 databases, and Tier 3 of 9 databases.

The database segmentation into tiers is solely based on the number of times they were cited in the SRoS, and not on the quality of the database. The quality of the database is a subjective measure, and it is not the focus of this work. Instead, the aim is to provide a detailed analysis of the databases referenced in the SRoS, and to propose a taxonomy with a minimum set of characterizing information that researchers should address when publishing or reviewing databases for AffC.

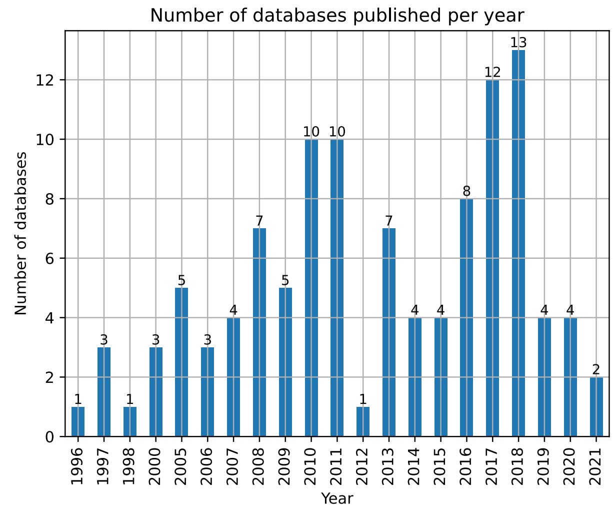

Fig. 1. Number of databases published per year.

TABLE I
Summary of the Number of References of Databases Tracked by Publication Year

Number of references Year	1	2	3	4	5	*Total*
1996	1	0	0	0	0	1
1997	2	2	0	0	0	4
1998	0	0	0	4	0	4
2000	2	2	0	0	0	4
2005	4	2	0	0	0	6
2006	0	2	3	4	0	9
2007	3	0	3	0	0	6
2008	2	10	0	0	0	12
2009	4	2	0	0	0	6
2010	4	4	6	0	10	24
2011	5	2	6	0	10	23
2012	0	2	0	0	0	2
2013	3	2	6	0	5	16
2014	1	6	0	0	0	7
2015	1	4	0	4	0	9
2016	4	8	0	0	0	12
2017	9	2	3	4	0	18
2018	9	6	3	0	0	18
2019	4	0	0	0	0	4
2020	4	0	0	0	0	4
2021	2	0	0	0	0	2
Total	64	56	30	16	25	191

The oldest database, Banse-Scherer [39], is from 1996, and the most recent are HEU [40] and VREED [41], from 2021. It was observed that 75% of the databases were created after 2009. Fig. 1 depicts the number of databases published per year, where it can be seen that the peak of database creation was in 2018 and 2017, with 13 and 12 databases, respectively. The lowest number of databases created was in 2012, with only one database.

In the following, the term "reference" is used for each database that was mentioned at least one time in each SRoS. It was observed that the 111 databases were referenced a total of 191 times in the SRoS. Table I summarizes the number of references of databases tracked by their publication year, and in Fig. 2 those values are plotted. Older databases have a chance of being referenced in more SRoS, as they have been around for a longer time, and have been used in more research works. Conversely,

TABLE II
STATISTICS ON THE RIGHTNESS OF THE URL, REFERENCE, YEAR, AND ACCESS AND AVAILABILITY STATUS OF THE SELECTED DATABASES

	Tier 1	Tier 2	Tier 3	All
URL‡	76.6 %	93.0 %	87.8 %	86.4 %
Ref†	15.6 %	34.9 %	34.1 %	28.3 %
Year*	34.4 %	48.8 %	70.7 %	48.7 %
Open datasets	53.1 %	73.7 %	100.0 %	64.0 %
LNA datasets	37.5 %	17.4 %	0.0 %	20.4 %

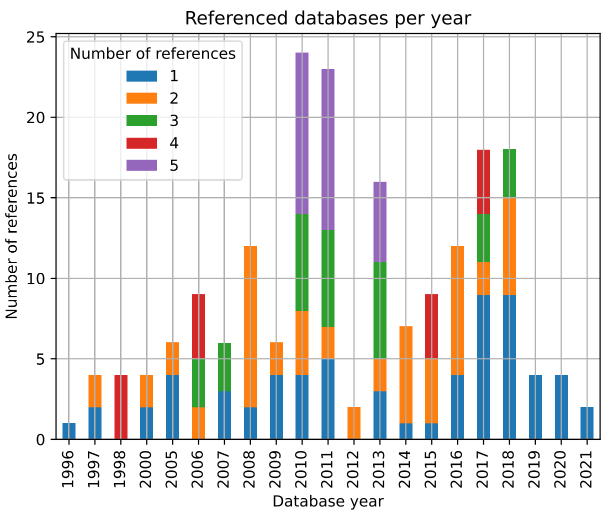

Fig. 2. References of databases tracked by publication year.

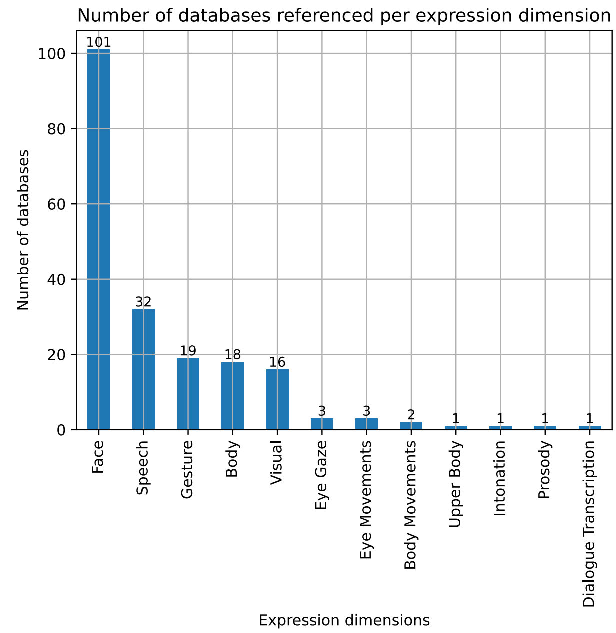

Fig. 3. Number of databases per expression dimension.

advances in technology and growing interest in the field of AffC, have led to the creation of new databases that are also being highly used in research works. The peak of references is for databases created in 2010 and 2011, with a total of 47 (24 + 23) references, and in 2017 and 2018 with a total of 36 (18 + 18) references. This is in line with the number of databases created in those years, as seen in Fig. 1, suggesting that those years are the most significant in terms of database creation and usage in AffC research works.

Modalities are a key factor in the selection of databases, as they can be used to categorize the databases into different groups. For instance, visual databases can be used for facial expression recognition, body databases for body gesture recognition, and physiological databases for sentiment analysis. Expression dimension refers to the modality features of the database, such as face, speech, gesture, body, visual, eye gaze, eye movement, etc. Fig. 3 shows the number of times that each expression dimension was mentioned in the 10 SRoS, as they were found in the SRoS papers. It should be noted that the figure translates the information obtained from the SRoS, which sometimes do not provide a clear distinction between the modalities. For instance, some databases were classified in the SRoS papers simply as "visual", not specifying if they were facial, body, or other types of visual databases. Analyzing Fig. 3, it is observed that visual databases are the most prevalent, followed by speech, and gesture. In more detail, facial databases were the most frequently referenced, with 101 references, showing the importance of facial expression recognition in the field of affective computing. Speech databases were the second most referenced, with 32 references, followed by gesture databases with 19 references.

In summary, from a macro level, as will be seen in the next sections, the databases are categorized as unimodal/multimodal (e.g., face, body, speech, physiological signals), data is acquired in controlled/uncontrolled environments (e.g., in-the-lab, in-the-wild), different emotion models are used (e.g., Ekman's [18] and Plutchik's [23] categorical models, and continuous dimensional models like [24], [25]), some are designed for sentiment analysis (negative, neutral, positive), and others give labels/targets for both emotion and sentiment analysis. It was observed that most visual databases (identified as face, gesture, body, visual etc. in the SRoS) are designed for emotion analysis, and most physiological databases are intended for sentiment analysis.

The databases are also annotated using different methods (e.g., self-report, expert annotator, crowdsourcing, etc.). A few databases take into account gender, age and cultural differences, which is important for model generalization. Next, a more detailed analysis of the database tiers is presented, which goes from the least to the most referenced databases, from Tier 1 to Tier 3.

A. Tier 1 Databases and Discussion

As mentioned in the previous section, Tier 1 consists of 64 databases that were referenced in only one of the selected SRoS. Table IV summarizes the Tier 1 databases.

The table is organized with the following columns: i) **Database**, which includes the name, year of publication, the original reference paper, access level (where P and O stands for "private" and "open", respectively), and the webpage of the database (where LNA stands for "link not available"); ii) **Static/dynamic** modality, based on the data provided on

TABLE III
EXAMPLES OF MINIMAL TAXONOMY IN AFFECTIVE COMPUTING DATABASES: CK+ (TIER 3), GEMEP (TIER 2), AND DFEW (TIER 1)

Characterization	**Information**		
Database version	Extended Cohn-Kanade (CK+) [49]	Geneva Multimodal Emotion Portrayal (GEMEP) [47]	Dynamic Facial Expression in-the-wild (DFEW) [60]
Modalities	Face	Face, body, gesture	Face
# Samples	593	1260	16372
# Subjects	123	10	*Not presented/Unclear*
(# Males, # Female)	*Not presented/Unclear*	[5, 5]	*Not presented/Unclear*
Subjects age	18-50	*Not presented/Unclear*	*Not presented/Unclear*
Ethnicity	Euro-American, Afro-American, other	*Not presented/Unclear*	*Not presented/Unclear*
Emotion model	Discrete: anger, contempt, disgust, fear, happiness, sadness, and surprise (+ Neutral)	Discrete: admiration, amusement, tenderness, hot anger (rage), disgust, despair, pride, anxiety, interest, irritation, elated joy, contempt, (panic) fear, pleasure, relief, sadness sand surprise (+ neutral)	Discrete: angry, disgust, fear, happy, sad and surprise (+ neutral)
Annotation method	Manual	Manual	Manual
Labelling approach	(i) manual Facial Action Coding System (FACS) coding on peak frames, (ii) apply label qualifying criteria (in terms of mapping facial Action Units (AUs) to emotions), and (iii) visual inspection of the clip from onset (neutral) to peak expression.	Pool of lay judges/raters.	Each sample was independently labelled by 10 well-trained annotators under professional guidance.
# Annotators per sample	15 % of the sequences were comparison coded by a second certified FACS coder	*Not presented/Unclear*	10
Environment	In-the-lab	In-the-lab	In-the-wild
Source type	Video	Audio and video	1500 videos/movies
Equipment	2 Panasonic AG-7500 Cameras	2 Sony DSR-PDX10 Cameras, and 3 microphones located at the cameras plus a Sennheiser headset microphone placed over the left ear of the actor	*Not presented/Unclear*
Ethical considerations	*Not presented/Unclear*	*Not presented/Unclear*	*Not presented/Unclear*
Resolution	640×490 and 640×480 pixels	720×576	*Not presented/Unclear*
Frame rate	30 frames per second	25 frames per second	*Not presented/Unclear*
File format	*Not presented/Unclear*	AVI, WAV	*Not presented/Unclear*
Pre-processing	*Not presented/Unclear*	Sound level normalization	*Not presented/Unclear*
Open/Private	Open	Open	Open
URL	https://www.jeffcohn.net/Resources/	https://www.unige.ch/cisa/gemep	https://dfew-dataset.github.io/download.html
License	Non-commercial academic research that is not subject to US export controls	Non-commercial academic research	Non-commercial academic research
Limitations and biases	Small number of samples, and lack of visual variability (in-the-lab vs in-the-wild)	Use of actors, and lack of visual variability (in-the-lab vs in-the-wild)	Class imbalance: angry (18.48%), disgust (1.22%), fear (8.14%), happy (20.63%), neutral (22.46%), sad (16.65%) and surprise (12.42%)

the SRoS that referenced the database, where pictures were considered as static, and videos, sequences of pictures, or any signal that changes over time as dynamic; iii) **SRoS referencing (Survey Ref.)** the database, where '†' indicates that the database was differently referenced in the SRoS (i.e., a different citation was used), '‡' indicates that the databases' URL was not given or is different from the one presented on the SRoS, and '*' means that the year, present in the first column, was corrected or added because it was not found in the reference or was an incorrect year of publication; iv) **Number of samples (#Samples)**, if present in the SRoS; v) **Number of subjects (#Subjects)**, if present in the SRoS; vi) **Number of emotions (#Emotions)**, if present in the SRoS, where 'd' indicates discrete and 'c' indicates continuous emotion models; vii) **Expression dimensions**, if present in the SRoS, which includes the following abbreviations, 'B' for 'body', 'F' for 'face', 'UB' for 'upper body', 'BM' for 'body movements', 'G' for 'gesture', 'S' for 'speech', 'I' for 'intonation', 'P' for 'prosody', 'DT' for 'dialogue transcription', 'EG' for 'eye gaze', 'EM' for 'eye movements', and 'Vi' for 'visual' (these abbreviations are common across other tier summary tables and are all listed here for clarity); viii) **Acted**, if present in the SRoS, where 'L' stands for "controlled / in-the-lab" and 'W' for "uncontrolled / in-the-wild".

Table IV reveals that many SRoS papers do not refer to the number of samples, which in the context of machine learning is a key factor, as the models need a considerable amount of data to be successfully trained. Most SRoS also do not refer, systematically, to the number and type of emotions that are present in the database, which is another key factor in selecting the database for a specific research objective.

Table II summarizes the statistics on the rightness of the data presented on the selected SRoS, including the URL, reference, year, access level, and availability status of the selected databases, and in Fig. 4 those values are plotted. In Tier 1, 64 databases were segmented, and from those 76.6% do not provide an access link to the database, or the provided link is no longer accessible (denoted by ‡ on the SRoS Ref. column of Table IV), 15.6% do not refer to the correct original reference paper of the database (denoted by † on the SRoS Ref. column), and 34.4% of the SRoS do not indicate or properly present the year of the database's original publication (denoted by '*' on the SRoS Ref. column).

Also, from Table IV, it is observed that 53.1% of the Tier 1 databases are publicly available. Note that it was considered as open or publicly available, not only databases that can be simply downloaded but also databases that require some form of authorization that is given to university staff members with permanent positions, as this is the case with the majority of the databases. To be clear, in this work, it was considered only as private, databases that are neither accessible to the public in

TABLE IV
TIER 1 DATABASES SUMMARY

Database	Static/Dynamic	Survey Ref.	#Samples	#Subjects	#Emotions	Expression dimensions	Acted
AFEW 7.0 2017 — [61] — P — LNA	D	[37]‡	1809		7d	F	(Posed, Spontaneous)
Affective Image Classification 2010 — [62] — O — https://tinyurl.com/2bdhb6mw	S	[17]					
Amazon 2007 — [63] — O — https://tinyurl.com/29vgsu8m	D	[17]	8000				
AVEC2013 2013 — [64] — P — LNA	D	[17]	340				
Banse-Scherer 1996 — [39] — P — LNA	D	[17]	224				
BAUM-2 2015 — [65] — O — https://tinyurl.com/2auvp2as	D	[11]‡	1047	286	7d	F, S	(Posed, Natural)
Belfast Natural database 2000 — [66] — P — LNA	D	[17]	239				(Acted, Natural)
BHUDES 2007 — [67] — P — LNA	D	[38]$^{\dagger *}$	323	5		S	(Acted)
BioVid Emo 2016 — [68] — O — https://tinyurl.com/28vp578k	D	[11]‡	430	94	5d		(Spontaneous, Induced)
Blogs 2009 — [69] — P — LNA	D	[17]*	252				
BP4D 2014 — [70] — P — https://tinyurl.com/2yobtqlx	S	[16]‡	328	41	7d	F	L
CALLAS 2010 — [71] — P — LNA	D	[11]$^{\dagger \ddagger *}$		21	3d	F, G, S	(Spontaneous, Induced)
CASIA 2017 — [72] — P — LNA	D	[38]		4	6d		(Acted)
CHEAVD 2.0 2018 — [73] — P — https://tinyurl.com/3lwmcs9	D	[11]‡	7030	527	8d		(Posed, Induced)
CLAS 2019 — [74] — O — https://tinyurl.com/2bk8tmb3	D	[11]‡		62			(Spontaneous, Induced)
DaFEx 2005 — [75] — P — LNA	D	[17]	1008				(Acted)
DFEW 2020 — [60] — O — https://tinyurl.com/2dyl3kuo	S	[16]‡	12059		7d	F	W
DREAMER 2017 — [76] — O — https://tinyurl.com/29xlabmv	S	[11]‡		23	2c		(Spontaneous, Induced)
DSdRD 2005 — [77] — P — LNA	D	[16]		24			
EMA 2017 — [78] — O — https://tinyurl.com/24xek4yt	D	[38]‡		3			(Acted)
EmoTV 2005 — [79] — P — LNA	D	[16]			14d	B, F, G, EG	
EU-Emotion Stimulus 2016 — [80] — P — LNA	D	[11]	418	19	21d	B, F, G, S	(Spontaneous, Induced)
FAU Aibo Emotion Corpus 2008 — [81] — P — https://tinyurl.com/2xpqch9f	D	[38]‡		51			(Natural)
GAPED 2011 — [82] — O — https://tinyurl.com/yqdzrdgg	S	[17]$^{\ddagger *}$	754				
Geneva airport lost luggage study 1997 — [83] — P — LNA	D	[17]*					(Natural)
HCR 2011 — [84] — P — LNA	D	[17]$^{\ddagger *}$					
HEU Emotion 2021 — [40] — P — LNA	D	[11]	19004	9951	10d	B, F, G, S	(Posed, Induced)
HOW 2011 — [85] — P — LNA	D	[16]‡				S, Vi	(Natural)
IMDB review 2011 — [86] — O — https://tinyurl.com/y2c5ee3j	D	[17]$^{\ddagger *}$	50000				
JACFEE 1997 — [87] — P — https://tinyurl.com/27az88ot	S	[17]$^{\ddagger *}$	56			F	
K-Emocon 2020 — [88] — O — https://tinyurl.com/2bzkzbzg	D	[12]$^{\dagger \ddagger *}$	16			F, G, S	(Spontaneous)
Keio-ESD 2017 — [89] — O — https://tinyurl.com/26k4t77v	D	[38]‡	940	71	47d		(Acted)
LDC 2018 — [90] — P — LNA	D	[38]	470	7	15d		(Acted)
MELD 2018 — [91] — O — https://tinyurl.com/2cd29cwo	D	[11]‡	13000		7d		(Posed, Induced)
MEmoR 2020 — [92] — O — https://tinyurl.com/223hhwoh	D	[12]$^{\dagger \ddagger *}$	5502				(Spontaneous)
MIT 2005 — [77] — O — https://tinyurl.com/2avlq5bw	D	[17]$^{\ddagger *}$					
MMSE/BP4D+ 2016 — [93] — O — https://tinyurl.com/2yobtqlx	D	[11]‡		140	10d		(Spontaneous, Induced)
MOUD 2013 — [94] — O — https://tinyurl.com/2yvuy49s	D	[17]$^{\ddagger *}$	80				
Movie Reviews 2020 — [95] — O — https://tinyurl.com/2c3coxwl	D	[17]$^{\ddagger *}$	2000				

TABLE IV
CONTINUE

Database	Static/Dynamic	Survey Ref.	#Samples	#Subjects	#Emotions	Expression dimensions	Acted
MPED 2019 — [96] — O — https://tinyurl.com/272xhv96	D	[11]‡		23	7d		(Spontaneous, Induced)
MUStARD 2019 — [97] — O — https://tinyurl.com/29trok9z	D	[11]‡	690		2d		(Posed, Induced)
NNIME 2017 — [98] — O — https://tinyurl.com/264wueuk	D	[11]‡		44	6d		(Posed, Natural)
NTUA 2017 — [98] — P — LNA	D	[11]†‡	204	22	2c	S	(Posed, Natural)
OMD 2009 — [99] — O — https://tinyurl.com/27dav8dm	D	[17]‡*	3269				
PMEmo 2018 — [100] — O — https://tinyurl.com/2xl4fu9y	D	[12]†‡*	457				(Posed)
RAMAS 2018 — [101] — P — LNA	D	[11]‡		10	6d	F, S	(Posed, Induced)
RAVDESS 2018 — [102] — O — https://tinyurl.com/298pt3eg	D	[11]‡	7356	24	8d		(Posed, Induced)
Reading/Leeds Emotional Speech Corpus 2000 — [103] — P — https://tinyurl.com/2xh7dxmd	D	[17]‡*					
RML 2008 — [104] — O — https://tinyurl.com/2a6y5kp5	D	[11]‡	720	8	6d	F, S	(Spontaneous, Natural)
SAL 2010 — [105] — P — LNA	D	[11]†‡	491	20	4d	F, S	(Spontaneous, Induced)
SAMM 2016 — [106] — O — https://tinyurl.com/2yy543tt	S	[16]‡*	159	32	>6d	F	L
SEED-IV 2018 — [107] — O — https://tinyurl.com/25enmrkp	D	[11]‡		15	4d	EM	(Spontaneous, Induced)
SEED-V 2019 — [108] — O — https://tinyurl.com/2da3vagl	D	[11]‡		16	5d	EM	(Spontaneous, Induced)
SEED-VIG 2017 — [109] — O — https://tinyurl.com/22mf8kne	S	[11]‡		23			(Spontaneous, Induced)
SEMAINE 2010 — [110] — O — https://tinyurl.com/23q9el34	D	[11]†‡*	959	150	5c	F, S	(Spontaneous, Induced)
SLADE 2018 — [111] — P — LNA	D	[11]‡			2c		(Spontaneous, Induced)
SMIC 2013 — [112] — P — https://tinyurl.com/2brb9ls5	S	[16]‡	164	16		F	L
Stanford Twitter Sentiment Test Set 2009 — [113] — P — LNA	D	[17]‡*					
TESS 2007 — [114] — O — https://tinyurl.com/w4rtmza	D	[38]‡	2800	2			(Acted)
TUM AVIC 2009 — [115] — P — LNA	D	[38]†‡*	3901	21			(Natural)
TURES 2017 — [116] — P — LNA	D	[38]‡	5100	582			(Acted)
VREED 2021 — [41] — O — https://tinyurl.com/29jy26ya	D	[12]†‡*	34			EG	(Posed)
WESAD 2018 — [117] — O — https://tinyurl.com/238pwm3b	D	[16]‡		15			
YouTube 2011 — [118] — O — https://tinyurl.com/28do8x5e	D	[17]‡*	47				

general nor to researchers through special academic licenses or authorizations. Following an in-depth examination of the URLs recorded in the table's respective field, it has come to light that 37.5% of the Tier 1 databases that were referenced are no longer reachable or accessible. This finding underscores the prevalence of unavailability among the listed databases, posing potential challenges to accessing the associated information or resources.

B. Tier 2 Databases and Discussion

Tier 2 databases consist of 38 databases that were referenced in 2 or 3 of the selected SRoS. Table V summarizes the Tier 2 databases, organized with the same columns as in Table IV, with the addition of the following ones: i) **Modalities**, where the extracted modalities from the SRoS are presented, which includes the following abbreviations, 'A' for 'audio', 'D' for 'depth', 'Im' for 'image', 'MC' for 'motion capture', 'NIR' for 'near infrared', 'Phy' for 'physiological', 'Psy' for 'psychophysiological', 'T' for 'text', 'Vd' for 'video', and '3D' for '3D AD'; and ii) **(#M,#F)**, where the number of subjects per gender ('M' stands for male and 'F' stands for female) is presented.

Modalities and gender information are relevant for researchers to select the database that best fits their research objectives. For instance, if a researcher is interested in facial expression recognition discriminating by gender, they should select databases that have gender information. If they are interested in multimodal emotion recognition, such as facial and speech, they should select databases that have multiple relevant modalities. These factors are not the only ones to be considered, as the number of samples, number of emotions, database availability, emotion model type, and other parameters are also key factors in selecting the database.

Since databases in this tier are referenced in two or three SRoS papers, when analyzing Table V, it is possible to observe

TABLE V
TIER 2 DATABASES SUMMARY

Database	Static/Dynamic	Survey Ref.	#Samples	#Subjects	#Emotions	Expression dimensions	Acted	Modalities	(#M,#F)
AffectNet 2017 — [42] — O — https://tinyurl.com/2ahjdpk8	S	[35]†‡*	440000		7d	F			
	S	[37]‡*	450000		7d	F	(Posed, Spontaneous)		
	S	[16]‡	450000		8d	F	W		
AMIGOS 2018 — [119] — O — https://tinyurl.com/275gqc2c	D	[34]‡*		40		Vi		A, D, Phy	
	D	[11]‡		40	7d+2c		(Spontaneous, Induced)	D, Phy, Vd	[27, 13]
	D	[16]‡*		40		B		Phy	
FABO 2006 — [120] — O — https://tinyurl.com/2d9s5uk7	D	[3]‡	206	23	10d	B, F, G, Vi	L		[11, 12]
	D	[11]‡		23	10d	B, F, G	(Posed, Induced)		[11, 12]
	D	[16]‡	1900	23	10d	B, F, G			
HUMAINE 2007 — [121] — O — LNA	D	[3]†‡	240	10	8d	B, F, G, Vi	L	A	[6, 4]
	D	[11]‡	50			F, G, S	(Posed, Natural)		
	D	[17]‡*	50				(Acted, Natural)	A, T, Vd	
Multi-PIE 2010 — [43] — P — https://tinyurl.com/29gm8gkl	S	[37]‡*	755370	337	6d	F	(Posed)		
	S	[35]†‡*	750000		7d	F			
	S	[36]‡		145	6d	F	L		
Oulu-CASIA 2011 — [44] — O — https://tinyurl.com/2a838guj	S	[35]†‡*	2880		6d	F			
	S	[37]‡*	2880	80	6d	F	(Posed)		
	S	[16]‡		80	7d	F	L		
RaFD 2010 — [122] — O — https://tinyurl.com/288v9mfj	S	[35]†‡*	8040		8d	F			
	S	[36]‡		67	8d	F	L		
	S	[37]‡*	1608	67	8d	F	(Posed)		
RECOLA 2013 — [123] — O — https://tinyurl.com/25akuzch	D	[11]‡		46	2c	F, S	(Spontaneous, Induced)		[19, 27]
	D	[16]‡		46		S, Vi	(Natural)	Phy	
	D	[38]†‡*		46			(Natural)	A, Vd	[19, 27]
SEED 2013 — [45] — O — https://tinyurl.com/29fjtcee	S	[12]†‡*	15		3d		(Posed)		
	S	[11]†‡*		15	3d		(Spontaneous, Induced)		[7, 8]
	D	[16]†*		15				Phy	
SFEW 2011 — [124] — O — https://tinyurl.com/2aq3qo2l	S	[35]†‡*	700		7d	F			
	S	[36]‡		95	7d	F	W		
	S	[11]‡	700	95	7d	F	(Spontaneous, Natural)		
4DFAB 2018 — [125] — P — https://tinyurl.com/24hrrck5	S	[37]‡*	1800000	180	7d	F	(Posed, Spontaneous)		
	S	[16]‡	1800000	180	6d	F	L		
AFEW 2012 — [126] — O — https://tinyurl.com/2aq3qo2l	S	[11]‡		330	6d	F	(Spontaneous, Natural)		
	D	[38]†‡*		330			(Natural)	A, Vd	
ASCERTAIN 2016 — [127] — O — https://tinyurl.com/2amneaq7	D	[34]‡*		58		F		Phy	
	D	[11]‡		58	5c	F	(Spontaneous, Induced)	Phy	[37, 21]
AVEC2014 2014 — [128] — P — LNA	D	[17]‡	300					A, Vd	
	D	[11]†‡*	340	292	3c	F, S	(Spontaneous, Induced)		
BAUM-1 2016 — [129] — O — https://tinyurl.com/26sebc9w	D	[11]‡	1222	31	6d	F, S	(Spontaneous, Induced)		[14, 17]
	D	[38]†‡*	1222	31			(Acted, Natural)	A, Vd	[18, 13]
BU-4DFE 2008 — [130] — P — https://tinyurl.com/2yobtqlx	S	[37]*	606	101	7d	F	(Posed)		
	S	[16]‡		101	7d	F	L		
CASME II 2014 — [131] — O — http://casme.psych.ac.cn/casme/e2	S	[35]†‡*	247			F			
	S	[16]‡	255	35	$>$6d	F	L		
CHEAVD 2017 — [132] — P — https://tinyurl.com/3lwmcs9	D	[11]‡		238	26d		(Posed, Induced)	A, Vd	[125, 113]
	D	[38]†‡*		238			(Acted natural)	A, Vd	
CK 2000 — [50] — O — https://tinyurl.com/25rkgh3o	S	[11]‡	486	97	7d	F	(Posed)		
	S	[16]‡		210	8d	F	L		
CMU-MOSEI 2018 — [133] — P — https://tinyurl.com/22q94gym	D	[11]‡	3228	1000	3d	F, G, S, I, P, Vi	(Spontaneous, Natural)	A	[570, 430]
	D	[16]‡				S, Vi	(Natural)	T	
CreativeIT 2010 — [134] — O — https://tinyurl.com/27k4c77n	D	[16]†		16		S, Vi	(Induced)	T	
	D	[11]†‡*	90	16	3c	B, BM, S	(Spontaneous, Induced)		[8, 8]
DES 1997 — [135] — P — LNA	D	[17]‡*						A	
	D	[38]†‡*		4			(Acted)	A	[2, 2]
EMILYA 2014 — [136] — P — LNA	D	[3]‡	7084	11	8d	B, G, Vi	L		[5, 6]
	D	[16]‡		11	8d	B, F, G			
Emo-DB 2005 — [137] — O — https://tinyurl.com/29wbf9ee	D	[11]‡	800	10	6d	S	(Posed, Natural)		[5, 5]
	D	[17]†‡*						A	
emoFBVP 2016 — [138] — O — LNA	D	[11]‡		10	23d	B, F, G, S	(Posed, Induced)	Phy	
	D	[12]†‡*	1380				(Spontaneous)	A, Phy, Vd	
EmotioNet 2016 — [139] — O — https://tinyurl.com/2yde53nl	S	[37]*	1000000		23d	F	(Posed, Spontaneous)		
	S	[16]‡				F	W		
eNTERFACE05 2006 — [140] — O — https://tinyurl.com/24hxrs8s	D	[11]‡	1166	42	6d	F, S	(Spontaneous, Induced)		[34, 8]
	D	[38]†‡*	1116	42			(Elicited)	A, Vd	[34, 8]
ExpW 2018 — [141] — O — https://tinyurl.com/2xmdufgg	S	[37]*	91793		7d	F	(Posed, Spontaneous)		
	S	[16]‡*			7d	F	W		
GEMEP 2010 — [47] — O — https://tinyurl.com/269kexxv	D	[3]†‡*	1260	10	18d	B, F, G, Vi	L	A	[5, 5]
	D	[16]‡		10	18d	B, F, G			
GEMEP-FERA 2011 — [46] — P — LNA	S	[35]†‡*	289		5d	F			
	D	[11]†‡	7000	10	18d	F, S	(Posed, Induced)		[5, 5]
ICT-MMO 2013 — [142] — O — https://tinyurl.com/23lgrusb	D	[17]‡*	370					A, T, Vd	
	D	[16]‡				S, Vi	(Natural)	T	
IEMOCAP 2008 — [143] — O — https://tinyurl.com/26j8px9d	D	[11]‡		10	9d	F, S, DT	(Spontaneous, Induced)		[5, 5]
	D	[16]		10		S, Vi	(Acted)	T	
KDEF 2008 — [144] — O — https://tinyurl.com/25fonqro	S	[36]‡		70	7d	F	L		
	S	[37]‡*	4900	70	7d	F	(Posed)		
LIRIS-ACCEDE 2015 — [145] — O — https://tinyurl.com/27ykpr5j	D	[3]†‡		64	6d	B, F, UB, G, Vi	L		[32, 32]
	D	[11]†‡*			2c		(Spontaneous, Natural)	A, Vd	
SAVEE 2008 — [48] — O — https://tinyurl.com/22k895me	D	[11]†‡*	480	4	7d	F, S	(Spontaneous, Induced)		[4, 0]
	D	[38]†‡*	120	14			(Acted)	A, Vd	[14, 0]
SFEW 2.0 2015 — [146] — O — https://tinyurl.com/228r3gea	S	[37]‡*	1776		7d	F	(Posed, Spontaneous)		
	S	[16]†‡			6d	F	W		
THEATER 2009 — [147] — P — LNA	D	[3]‡	258		8d	B, G, Vi	W	A	
	D	[16]‡			8d	B, G			
VAM 2008 — [148] — O — https://tinyurl.com/2cnd3xe7	D	[11]‡		104	3c	F, S	(Spontaneous, Natural)		
	D	[38]†‡*		47	3c		(Natural)	A, Vd	

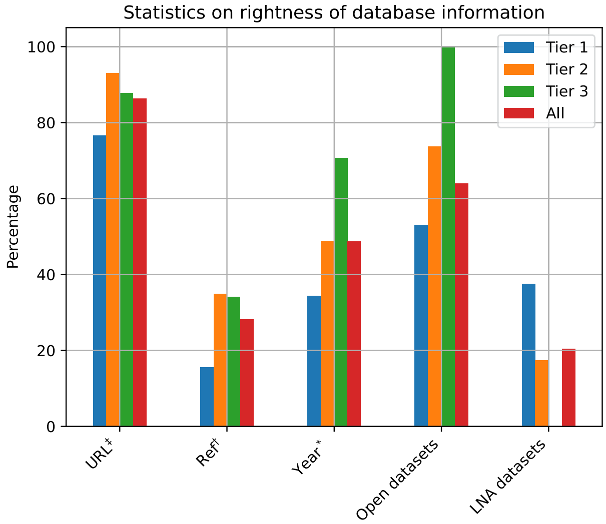

Fig. 4. Statistics on the rightness of the URL, reference, year, access, and availability status of the selected databases.

that for the same database, different data is presented by different SRoS authors. Some cases even present conflicting data, like in the AffectNet database [42], where Baffour et al. [35] indicates that the database has 7 discrete emotion categories and Wang et al. [16] indicates that it has 8 (according to the original reference paper, the AffectNet database has 8 emotion category labels [42]). The same can be observed in the Multi-PIE database [43], where Baffour et al. [35] reports that the database has 6 discrete emotion categories, and Jampour and Javidi [36] report that it has 7 (according to the reference database paper it has 6 emotion categories [43]).

Another example is found in the Oulu-CASIA database [44], where Baffour et al. [35] reports that the database has 6 discrete emotion categories and Wang et al. [16] reports that the database has 7 (the reference paper indicates that there are 6 emotion categories on this database [44]). Other discrepancies occur in the number of samples, acted, modalities etc. For instance, depending on the SRoS, it was possible to conclude that the SEED database [45] is static [11], [12] or dynamic [16]. This clearly shows that there is inconsistency in the information provided by different SRoS authors, which may or may not be relevant in certain cases.

It is also worth noting that the number of samples and subjects differs between different SRoS papers. E.g., for the GEMEP-FERA database [46], Siddiqui et al. [11] mentions 7000 samples, while Baffour et al. [35] mentions 289 samples. The reference paper of the database mentions that the entire corpus of the GEMEP database [47] is over 7000 samples, however the GEMEP-FERA emotion sub-challenge consists of a subset of 289 samples [46]. For the SAVEE database [48], Siddiqui et al. [11] mentions the number of samples as 480, and the number of subjects as 4, while Yadav et al. [38] mentions 120 samples, and 14 subjects. According to both the database reference paper and the companion website, the number of samples is 480, and the number of subjects is 4 [48].

In general, the same conclusions regarding the publication year, database reference publication, and URL, can be drawn for the analysis of the Tier 2 databases, as from the analysis of the Tier 1 databases (see Table II). Specifically, from the 38 databases segmented in Tier 2, 48.8% do not properly indicate (is absent or wrong) the year of the database's original publication, 34.9% do not refer to the correct original reference paper of the database, and 93.0% do not give an access link to the database or the provided link is no longer accessible. It is also observed that 73.7% of the Tier 2 databases are publicly available (see criteria to consider a database as open or private in the prior Section). This suggests that databases which are referenced more frequently are more likely to be publicly accessible. Finally, after examining the URLs presented in Table V, it has come to light that 17.4% of the databases referenced are no longer reachable or accessible.

Some partial conclusions can be made from Tier 2 databases. For instance, some cases show a noted inconsistency in the information provided by different SRoS authors regarding the same databases. This inconsistency may lead to confusion and incorrect conclusions when selecting a database for research. As an example, discrepancies are observed in the number of emotion categories and other database characteristics reported by different studies. Apart from modalities and gender information, many other factors like the number of samples, number of subjects, the range and type of emotions, and database availability, vital in selecting a database for research, are absent in the SRoS papers.

Public accessibility of databases indicates a correlation between the frequency of references and public accessibility, i.e., databases that are more frequently referenced tend to be more likely to be publicly accessible, suggesting that visibility through references may influence or correlate with accessibility. A considerable number of the databases suffer from issues such as incorrect or missing publication years, erroneous reference papers, and inaccessible URLs. There is a notable unavailability issue with the databases, highlighting a broader issue of persistence and accessibility in academic databases, which may restrict the reproducibility and replicability of research findings.

C. *Tier 3 Databases and Discussion*

Tier 3 databases consist of 9 databases that were referenced in at least 4 SRoS papers. Table VI summarizes the Tier 3 databases collected information. The table is organized with the same columns as Table V, with the addition of the following ones: i) **Emotions**, if present in the SRoS, which includes the following abbreviations, 'Amu' for 'Amusement', 'Ang' for 'Anger', 'Anx' for 'Anxiety', 'Con' for 'Contempt', 'Dis' for 'Disgust', 'Exc' for 'Exciting', 'Fea' for 'Fear', 'Fun' for 'Funny', 'Hap' for 'Happiness', 'Joy' for 'Joy', 'Neu' for 'Neutral', 'Sad' for 'Sadness', 'Sho' for 'Shock', and 'Sur' for 'Surprise'; and ii) **Physiological signals**, if present in the SRoS, using the following abbreviations, 'B-hEOG' for 'bipolar horizontal electrooculogram', 'B-tEMG' for 'bipolar trapezius electromyography', 'BS' for 'brain signals', 'ECG' for 'electrocardiogram', 'EEG' for 'electroencephalogram', 'EMG' for 'electromyography', 'EOG' for 'electrooculography', 'GSR' for 'galvanic skin response', 'MEG' for 'magnetoencephalography', 'RESP' for 'respiration', and 'TEMP' for 'temperature'. Additionally, for each database, is included an authors'

TABLE VI
TIER 3 DATABASES SUMMARY

Database	Static/Dynamic	Survey Ref.	#Samples	#Subjects	#Emotions	Expression dimensions	Acted	Modalities	(#M,#F)	Emotions	Physiological signals
CK+ 2010 — [49] — O https://tinyurl.com/25rkgh3o	S	[35]†‡*	593	123	7d	F				Ang,Con,Dis,Fea,Hap,Sad,Sur	
	S	[37]‡*	593	123	8d	F	(Posed, Spontaneous)				
	S	[11]‡	593	123	7d	F	(Posed and spontaneous, Natural)				
	S	[16]‡	593	123	8d	F	L				
	D	[17]‡*		97				Vd			
	D	**Ours**	**593**	**123**	**7d+1n**	**F**	**L(Posed and spontaneous, Natural)**	**Vd**		**Ang,Con,Dis,Fea,Hap,Sad,Sur**	
DEAP 2011 — [51] — O https://tinyurl.com/2xm6syf8	D	[17]*						Phy, Vd			
	D	[12]†‡*	40			F	(Spontaneous)	Phy, Vd			
	D	[34]‡*		32				Phy			
	D	[11]‡*		32	4c	F	(Spontaneous, Induced)	Phy			EEG
	D	[16]*		32				Phy			EEG,EMG,EOG,GSR,RESP
	D	**Ours**	**1280**	**32**	**4c (valence, arousal, dominance, liking)**	**F**	**L(Spontaneous, Induced)**	**Phy, Vd**	**[16, 16]**		**EEG,EMG,EOG,GSR,RESP,TEMP**
FER2013 2013 — [52] — O https://tinyurl.com/2jev5wa7	S	[37]‡*	35887		7d	F	(Posed, Spontaneous)				
	S	[12]†‡*	35887				(Posed)	Im			
	S	[35]†‡*	35887			F				Ang,Dis,Fea,Hap,Sad,Sur,Neu	
	S	[11]†‡*	35887		7d	F	(Spontaneous, Natural)				
	S	[16]‡	35887		7d	F	W				
	S	**Ours**	**35887**		**7d**	**F**	**W**	**Im**		**Ang,Dis,Fea,Hap,Sad,Sur,Neu**	
MAHNOB-HCI 2011 — [53] — O https://tinyurl.com/2xrm75bw	S	[12]†‡*					(Posed)				ECG,EEG,GSR,RESP,TEMP
	D	[17]†‡*						A, Phy, Vd			
	D	[11]‡		27	3c	B, F, BM, EM	(Spontaneous, Induced)	A, Phy	[11, 16]		EEG
	D	[16]‡*		27		S, Vi	(Induced)	Phy		Amu,Dis,Fea,Joy,Sad,Neu	
	D	[34]‡*		27		F, EG		A, Phy			
	D	**Ours**	**1674**	**27**	**9d**	**F, UB, BM, EG**	**L(Spontaneous, Induced)**	**A, Phy, Vd**	**[11, 16]**	**Amu,Ang,Anx,Dis,Fea,Joy,Sad,Sur,Neu**	**ECG,EEG,GSR,RESP,TEMP**
MMI 2010 — [55] — O https://tinyurl.com/2ykuoxy4	S	[35]†‡*	2900		7d	F				Ang,Dis,Fea,Hap,Sad,Sur,Neu	
	S	[37]*		25	7d	F	(Posed)				
	S	[11]†‡*		94		F	(Posed, Natural)				
	S	[16]‡	2900	25	7d	F	L				
	D	[17]†‡*	2900			F		Vd			
	D	**Ours**	**2900**	**75**	**6d+1n**	**F**	**L(Spontaneous, Induced)**	**Im, Vd**		**Ang,Dis,Fea,Hap,Sad,Sur,Neu**	
BU-3DFE 2006 — [56] — O https://tinyurl.com/2yobtqlx	S	[35]†‡*	2500		7d	F				Ang,Dis,Fea,Hap,Sad,Sur,Neu	
	S	[36]‡		100	7d	F	L			Ang,Dis,Fea,Hap,Sad,Sur,Neu	
	S	[37]*	2500	100	7d	F	(Posed)				
	S	[16]‡	2500	100	7d	F	L				
	S	**Ours**	**5000**	**100**	**7d**	**F**	**L**	**Im**	**[44, 56]**	**Ang,Dis,Fea,Hap,Sad,Sur,Neu**	
DECAF 2015 — [57] — O https://tinyurl.com/25yycpem	D	[34]‡*		30		F		Phy			
	D	[17]‡*						Phy, Vd			
	D	[11]‡		30	2c	F	(Spontaneous, Induced)	NIR, Phy	[16, 14]		B-hEOG,B-tEMG,BS,ECG,MEG
	D	[16]‡		30		S, Vi	(Induced)	Phy		Amu,Exc,Fun	
	D	**Ours**	**76**	**30**	**9d + 3c (valence, arousal, dominance)**	**F**	**L**	**NIR, Phy, Vd**	**[16, 14]**	**Amu,Ang,Dis,Exc,Fea,Fun,Hap,Sad,Sho**	**B-hEOG,B-tEMG,BS,ECG,MEG**
JAFFE 1998 — [58] — O https://tinyurl.com/24waqc4o	S	[35]†‡*	213		7d	F			['-', 213]	Ang,Dis,Fea,Hap,Sad,Sur,Neu	
	S	[37]‡*		10	7d	F	(Posed)				
	S	[11]†‡*		10	7d	F	(Posed, Natural)				
	S	[16]‡		10	7d	F	L				
	S	**Ours**	**213**	**10**	**7d**	**F**	**L**	**Im**	**['-', 10]**	**Ang,Dis,Fea,Hap,Sad,Sur,Neu**	
RAF-DB 2017 — [59] — O https://tinyurl.com/22p7hxja	S	[36]‡	29672		7d	F	W			Ang,Dis,Fea,Hap,Sad,Sur,Neu	
	S	[37]*	29672		7d+12d	F	(Posed, Spontaneous)				
	S	[16]‡	29672		7d	F	W				
	S	[35]†‡*	30000		7d	F				Ang,Dis,Fea,Hap,Sad,Sur,Neu	
	S	**Ours**	**29672**		**7d**	**F**	**W**	**Im**		**Ang,Dis,Fea,Hap,Sad,Sur,Neu**	

row (identified as "Ours") where the information extracted from the original reference paper/website of the database is presented.

Since Tier 3 databases are the most referenced within the selected SRoS papers, it is expected that the information is more consistent among different SRoS authors, as well as the completeness and rightness of the information that is provided, given that they are the databases with the highest visibility. However, it is still observed that from the 9 databases segmented in Tier 3, 70.7% of the SRoS authors do not properly indicate (or indicate at all) the year of the database's original publication, 34.1% do not refer to the correct original reference paper of the database, and 87.8% do not give an access link to the database, or the link provided is no longer accessible. It is also observed that 100.0% of the databases are publicly available (see public/private criteria on Section III-A). Finally, after examining the URLs presented in Table VI, it has come to light that all databases are reachable or accessible. It is also worth noticing that these databases are all from before 2018, indicating that database usage by the research community might take some time to be consolidated. Adoption should not only be related to the novelty of the database but also to the quality of the data and the relevance of the database to the research objectives. Further, some databases are very specific and hard to collect, like the ones with physiological signals, which may also explain the higher number of references in the SRoS papers.

Looking at Table VI, it is observed that it is composed of 5 databases with 5 references each on the selected SRoS papers, and 4 databases with 4 references each. Next, a more detailed analysis of the Tier 3 databases is given.

The Extended Cohn-Kanade (CK+) database [49], is an extension of the original Cohn-Kanade (CK) database [50] (present in Tier 2), designed for research on facial expression classification. The database comprises 593 video sequences from 123 different subjects, with ages spanning from 18 to 50 years old. The videos capture facial shifts from neutral expressions to targeted peak expressions. Among these sequences, 327 are labeled with one of seven expression classes, namely: anger, contempt, disgust, fear, happiness, sadness, and surprise. The videos were recorded at a frame rate of 30 frames per second (FPS) and have resolutions of either 640 × 490 or 640 × 480 pixels. CK+ is widely used as a benchmark for developing and evaluating facial expression classification algorithms and is one of the most extensively used laboratory-controlled facial expression databases.

The Dataset for Emotion Analysis using Physiological and Audiovisual Signals (DEAP) [51] is a multimodal dataset, designed for the analysis of human affective states. It includes the electroencephalogram (EEG) and peripheral physiological signals of 32 participants, recorded as they watched 40 1-minute-long excerpts of music videos. Participants rated each video on arousal, valence, like/dislike, dominance, and familiarity. For 22 of these 32 participants, a frontal face video was also recorded. The dataset was collected using a novel method for stimuli selection, which involved retrieval by affective tags from the *last.fm* website, video highlight detection, and an online assessment tool.

The Facial Expression Recognition 2013 (FER2013) database [52] is widely used for facial expression recognition. It contains 35,887 facial RGB images, each of size 48 $\times$ 48 pixels. These images are labeled into seven categories, namely: angry, disgust, fear, happy, sad, surprise, and neutral. Among these, the disgust category has the fewest images, 547, while the other categories have 4,953 for Anger, 5,121 for Fear, 8,989 for Happiness, 6,077 for Sadness, 4,002 for Surprise, and 6,198 for Neutral, showing a significant class imbalance.

MAHNOB-HCI [53] is a multimodal database designed for emotion recognition and implicit tagging research. It contains synchronized recordings from six camera views, a head-worn microphone, a room microphone, an eye gaze tracker, and physiological sensors measuring ECG, 32-channel EEG, respiration pattern, GSR, and skin temperature, to capture spontaneous emotional responses. The data was collected from participants as they interacted with various multimedia content, including movie clips and images. In addition to their emotional reactions, the database also records the participant's agreement or disagreement with the applicability of tags presented alongside the media items. The database uses 9 emotional keywords assigned to all videos, which are amusement, anger, anxiety, disgust, fear, joy, neutral, sadness, and surprise. The emotional keywords are then used to map the emotional responses of the participants to valence-arousal classes as detailed in Fontaine et al. [54]. The database is composed of 27 subjects, of which 11 are males and 16 are females.

The MultiModal Interaction (MMI) facial expression database [55] was introduced in 2002 aiming to provide large volumes of visual data for facial expression analysis. From [55], it was not possible to find clear information on the total number of samples and the number of males and females. However, the companion website states that the database is composed of 75 subjects and that it contains over 2,900 videos and high-resolution still images, capturing the full temporal pattern of facial expressions, from neutral to onset, apex, and offset phases, and back to a neutral face. The data was collected "in-the-lab", and it includes both prototypical expressions of the six basic emotions and expressions with a single facial action coding system (FACS) action unit (AU) activated, covering existing AUs and many other action descriptors.

Since its introduction, the database has been expanded to include recordings of naturalistic expressions. It has been annotated to detect AUs within videos, with partial coding at the frame level. This coding indicates, for each frame, whether an AU is present in its neutral, onset, apex, or offset phase. The MMI Facial Expression Database is publicly available to the scientific community, and the URL is accessible.

The Binghamton University 3D Facial Expression (BU-3DFE) database [56] is used for 3D facial expression recognition research tasks, using static data. It consists of 2,500 facial expression models, each derived from one of 100 subjects that span a wide range of ethnic backgrounds. The subjects were 56% female and 44% male, with ages ranging from 18 to 70 years old. Each subject has performed seven distinct expressions, namely neutral, happiness, disgust, fear, anger, surprise, and sadness. Each of these expressions, excluding the neutral one, was represented at four different intensity levels. This results in a total of 25 unique 3D expression models per subject. Alongside each 3D expression model, the database provides corresponding facial texture images captured from two different views, approximately $\pm 45^{\circ}$. Consequently, the database contains 2,500 texture images from two viewpoints and 2,500 geometric shape models.

The multimodal dataset for DECoding user physiological responses to AFfective multimedia content (DECAF) [57] is a multimodal database designed for decoding and understanding user physiological responses to affective multimedia content. It contains brain signals acquired using a magnetoencephalogram (MEG) sensor, which allows for minimal physical contact with the user's scalp, thereby facilitating a more naturalistic affective response. It includes the emotional responses of 30 participants to 40 1-minute music video segments (the same ones used in DEAP database) and 36 movie clips. This allows for a comparative analysis between the EEG and MEG modalities, as well as between movie and music stimuli for affect recognition. The database also contains synchronously recorded near-infrared (NIR) facial videos, horizontal Electrooculogram (hEOG), Electrocardiogram (ECG), and trapezius-Electromyogram (tEMG) peripheral physiological responses.

Additionally, the database includes continuous emotion annotations over time for movie clips sourced from seven users. These annotations are utilized to demonstrate dynamic emotion prediction capabilities. It has been used for analyzing correlations between participant self-assessments and their physiological responses, and for single-trial classification results for valence, arousal, and dominance, providing valuable insights into the complex interplay between physiological responses and emotional states.

The Japanese Female Facial Expression (JAFFE) database [58] is used for facial expression recognition. It consists of 213 images, with a resolution of 256 $\times$ 256 pixels, of different posed facial expressions, from 10 different Japanese female subjects. Each subject was asked to perform 7 facial expressions, specifically the 6 basic facial expressions (happiness, sadness, surprise, anger, disgust, and fear) and a neutral expression. These images were annotated with average semantic ratings on each facial expression by 60 annotators.

The Real-world Affective Faces Database (RAF-DB) [59] is a comprehensive facial expression database consisting of 29,672 images, each independently annotated with basic or compound expressions by 40 annotators. The images in the database are diverse, featuring subjects of varying age, gender, and ethnicity, and captured under different lighting conditions and head poses. Some images include occlusions like glasses and facial hair.

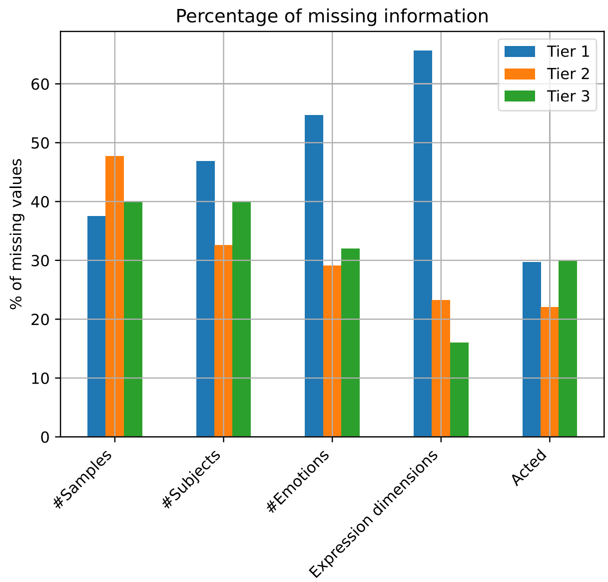

Fig. 5. Percentage of missing information, by tier, in the selected SRoS papers.

Each image in RAF-DB is accompanied by a wealth of annotations, including a 7-dimensional expression distribution vector, accurate and automatic landmark locations, bounding boxes, and attributes such as race, age range, and gender. Included is a single-label subset with 7 basic emotions (surprised, fearful, disgusted, happy, sad, angry, and neutral) and a two-tab subset with 12 compound emotions (fearfully surprised, sadly angry, sadly fearful, angrily disgusted, angrily surprised, sadly disgusted, fearfully disgusted, disgustedly surprised, happily surprised, sadly surprised, fearfully angry, and happily disgusted). In the next Section, a taxonomy for characterizing databases in AffC is proposed.

IV. Database Proposed Taxonomy

A SRoS is a valuable tool for researchers to understand the current state-of-the-art, identify gaps, and propose future research directions. It should also provide comprehensive information about the field it covers, allowing researchers to easily identify the most relevant and useful information for their research objectives. Therefore, the completeness and rightness of the information provided are essential for the SRoS to be useful and relevant.

However, from the analysis of the selected SRoS papers, it is observed that the information that is provided by the SRoS authors does not follow a common standard and, in some cases, is not always consistent, complete, or correct. Fig. 5 shows the percentage of missing information, by tier, in the selected SRoS papers. It is possible to observe that basic characteristics like the number of samples and subjects, with missing percentages above 30%, are most of the time neglected. Also, visible is that the number of emotions, expression dimensions, and acted information is also frequently missing. In general, missing information is clearly higher in Tier 1 databases, showing a tendency for the more referenced databases to be likely more complete and have more consistent information.

Fields like modalities, number of samples, number and type of emotions, age of the subjects, ethnicity, environments etc. allow a fast and easy selection of the database that best fits the research objectives. Researchers need clear information about which databases contain full body emotion data if they are interested in developing models for body emotion recognition. Similarly, it should also be clear which databases include the subject's age, gender, ethnicity, and other relevant data if they want to develop models that make these distinctions.

In this context, a taxonomy is proposed, with a minimum set of information, to be followed when publishing or reviewing databases (many times it was very difficult or even impossible to find relevant information such as in the cases mentioned above) or by SRoS authors. Essential elements that should be clearly presented include the following. Present the (i) modality and data type by indicating if the database consists of text, picture, video, audio, physiological signals (such as EEG and ECG), or a mix of these, and describe the integration of each modality if it is multimodal. Another relevant aspect is the (ii) diversity and size of data. In these cases, authors should describe the quantity of samples, number of subjects, data points in the database, and explain the variety in terms of scenarios and demographics (e.g., age, gender, and ethnicity). (iii) Labelling and annotation information that describe the different emotional models that are used (e.g., dimensional models, like, arousal-valence; discrete emotions, like, happy, or sad), and also the number of annotators per sample, as well as the methods used to annotate the data (manual annotation, crowdsourcing, automated tools etc.). If possible, inform on how consistently raters annotate various data. The (iv) source and method of collection informing where the data came from (e.g., public domain, online platforms, controlled experiments etc.), its nature (e.g, equipment that was used and the environment), since this is important to understand the context in which the data was collected, and to assess its relevance and usefulness for a specific application. (v) Ethical considerations and privacy issues are also important to be addressed. The authors should explain how the data was collected, how the consent of the subjects was acquired, and how their privacy is safeguarded. This includes data anonymization and the protection of sensitive information, such as personal data and health information.

Supply information on the (vi) data quality, including resolution, frame rate, file format, and other relevant parameters for both visual and audio data. In the case of physiological signals, the authors should provide information, for example, on the sampling rate, the number of channels, and the type of signals that were recorded. Here is also important to mention which pre-processing techniques were used on the data, such as noise reduction and normalization. The (vii) availability and accessibility of the database is also an important aspect to be addressed. As such, the authors should provide details about obtaining the database (publicly accessible, limited usage, upon request etc.), and indicate the conditions of the database license. The (viii) limitations and biases of the database should also be addressed. This includes any biases or limitations that are known to exist in the database (such as cultural biases or the over-representation of particular emotions). Finally, the (ix) version and modifications of the database should be addressed. The authors should give specifics about the database version and any modifications or additions that have been made over time, if any.

Other aspects associated with the type of methods used with the databases, such as the type of features used, the type of classifiers, the type of pre-processing, etc., can also be addressed. This provides readers the capacity to identify adequate databases for their research and application development and to understand the relevance and usefulness of the database in the context of AffC. For instance, this might be useful for researchers who are looking for a database that is representative of a specific demographic or emotional state, or for those who are looking for a database that is suitable for a specific type of analysis or application.

Table III shows an example of how the proposed taxonomy can be applied to the CK+, GEMEP, and DFEW databases [47], [49], [60]. These databases were chosen as an example, each belonging to a different tier, and due to their prominence in the field of AffC and their diverse characteristics. The data on the table was collected from the original reference papers of the databases, and shows that the database authors did not provide (or it is not clear) all the information that is proposed in the taxonomy, like the number of subjects, number of males and females, age range, file format, pre-processing techniques, and ethical considerations. This shows the value of the proposed taxonomy to provide a standard for presenting databases, ensuring that essential information is included, and that the information is consistent and complete.

V. Conclusion

In this paper, an in-depth analysis of SRoS papers on AffC databases is conducted, where the main objective is to identify how AffC databases are presented in the literature, as well as their relevance. This was done by selecting 10 SRoS from a more general list of SRoS papers. The papers were selected based on the criteria detailed in Section II-A, namely the paper must be a systematic review or survey paper, must be published in a journal, must be written in English, must be available online, and must be published between 2020 and February 2024. From the analysis, it was found that a variety of methodologies were used to present information about AffC databases. Tables IV, V, and VI summarized the databases that were referenced in the selected SRoS papers, and the information that was provided by the SRoS authors. From the 111 databases that were referenced in the selected SRoS, 64 were referenced in only one SRoS, 38 were referenced in 2 or 3 SRoS, and 9 were referenced in at least 4 SRoS. The databases were segmented into three tiers, and the information that was provided by the SRoS authors was analyzed.

The analysis revealed inconsistencies and discrepancies in the details presented across different SRoS. Additionally, the information provided by SRoS authors was not always complete, with missing information in certain cases.

To address these challenges, a minimal taxonomy for publishing or reviewing affective computing databases was proposed, which includes a set of essential elements that should be clearly presented in the SRoS papers that address the subject. These elements include the modality and data type, diversity and size of data, labeling and annotation information, source and method of collection, ethical considerations and privacy issues, data quality, availability and accessibility of the database, limitations and biases of the database, and version and modifications of the database. This taxonomy will help researchers to easily identify the most relevant and useful information for their research objectives, and to select the database that best fits their research objectives.

Summarizing, the primary goal of this study was to provide a systematic review of affective computing databases referenced in recent SRoS, highlighting their relative importance and utility, being the main contributions: i) a comprehensive tier-based database classification – unlike previous surveys, a tiered classification system was introduced based on how frequently the databases were referenced across the selected SRoS. This provides a practical tool for researchers to quickly identify the most frequently cited and, perhaps more controversially, more impactful databases; ii) Identifying data discrepancies – an in-depth analysis uncovered discrepancies in the information provided across different SRoS. It was demonstrated that crucial information (such as number of samples, number of emotions, and modality) is often inconsistent or incomplete. This is a key finding not emphasized in previous surveys, highlighting the need for standardized reporting; iii) Database taxonomy proposal – a minimal taxonomy for reporting affective computing databases was proposed, designed to improve the clarity and utility of future SRoS. This taxonomy suggests a standard for presenting databases, ensuring that essential information such as demographic details, modalities, annotation method, and labelling approach are included.

Based on this research, actionable recommendations are provided for future databases and SRoS to improve the quality and consistency of data reporting. These insights have not been explicitly discussed in previous works and can guide to improvements in SRoS practices. While this survey does not delve deeply into database-derived results, gaps in the literature related to database characterization, transparency, and consistency were addressed.

Future work consists of analyzing, following the same principle, the models and methods available for emotion and sentiment analysis.

References

[1] R. Z. Cabada, H. M. C. López, and H. J. Escalante, *Multimodal Affective Computing: Technologies and Applications in Learning Environments*. Berlin, Germany: Springer, 2023.

[2] G. K. Verma, *Multimodal Affective Computing: Affective Information Representation, Modelling, and Analysis*. Bentham Science Publishers, Mar. 2023, doi: 10.2174/97898151244531230101.

[3] F. Noroozi, C. A. Corneanu, D. Kaminska, T. Sapinski, S. Escalera, and G. Anbarjafari, "Survey on emotional body gesture recognition," *IEEE Trans. Affect. Comput.*, vol. 12, no. 2, pp. 505–523, Second Quarter, 2021.

[4] S. Turaev et al., "Data-driven analysis of patients' body language in healthcare: A comprehensive review," *IEEE Access*, vol. 12, pp. 16514–16548, 2024.

[5] M.-A. Mahfoudi, A. Meyer, T. Gaudin, A. Buendia, and S. Bouakaz, "Emotion expression in human body posture and movement: A survey on intelligible motion factors, quantification and validation," *IEEE Trans. Affective Comput.*, vol. 14, no. 4, pp. 2697–2721, Fourth Quarter, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/9968283/

[6] S. G. Tejashwini and D. Aradhana, "Multimodal deep learning approach for real-time sentiment analysis in video streaming," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, 2023, Art. no. 2023. [Online]. Available: http://thesai.org/Publications/ViewPaper?Volume=14&Issue=8&Code=IJACSA&SerialNo=81

[7] Y. Gao, Y. Fu, M. Sun, and F. Gao, "Multi-modal hierarchical empathetic framework for social robots with affective body control," *IEEE Trans. Affect. Comput.*, vol. 15, no. 3, pp. 1621–1633, Third Quarter, 2024.

[8] P. J. Cardoso, J. M. Rodrigues, and R. Novais, "Multimodal emotion classification supported in the aggregation of pre-trained classification models," in *Proc. Int. Conf. Comput. Sci.*, 2023, pp. 433–447.
[9] T.-P. Jung et al., "Utilizing deep learning towards multi-modal biosensing and vision-based affective computing," *IEEE Trans. Affect. Comput.*, vol. 13, no. 1, pp. 96–107, First Quarter, 2022.
[10] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," *Inf. Fusion*, vol. 37, pp. 98–125, 2017.
[11] M. F. H. Siddiqui, P. Dhakal, X. Yang, and A. Y. Javaid, "A survey on databases for multimodal emotion recognition and an introduction to the VIRI (visible and infrared image) database," *Multimodal Technol. Interact.*, vol. 6, no. 6, Jun. 2022, doi: 10.3390/mti6060047.
[12] N. Ahmed, Z. Al Aghbari, and S. Girija, "A systematic survey on multimodal emotion recognition using learning algorithms," *Intell. Syst. Appl.*, vol. 17, 2023, Art. no. 200171.
[13] L. Shu et al., "A review of emotion recognition using physiological signals," *Sensors*, vol. 18, no. 7, 2018, Art. no. 2074.
[14] H. Gunes and M. Piccardi, "Bi-modal emotion recognition from expressive face and body gestures," *J. Netw. Comput. Appl.*, vol. 30, no. 4, pp. 1334–1345, 2007.
[15] M. Munezero, C. S. Montero, E. Sutinen, and J. Pajunen, "Are they different? affect, feeling, emotion, sentiment, and opinion detection in text," *IEEE Trans. Affect. Comput.*, vol. 5, no. 2, pp. 101–111, Second Quarter, 2014.
[16] Y. Wang et al., "A systematic review on affective computing: Emotion models, databases, and recent advances," *Inf. Fusion*, vol. 83/84, pp. 19–52, 2022.
[17] N. J. Shoumy, L. M. Ang, K. P. Seng, D. M. Rahaman, and T. Zia, "Multimodal big data affective analytics: A comprehensive survey using text, audio, visual and physiological signals," *J. Netw. Comput. Appl.*, vol. 149, Jan. 2020, doi: 10.1016/j.jnca.2019.102447.
[18] P. Ekman, "Universals and cultural differences in facial expressions of emotion," in *Nebraska symposium on Motivation*, Lincoln, NE, USA: Univ. Nebraska Press, 1971.
[19] P. Ekman and W. V. Friesen, "A new pan-cultural facial expression of emotion," *Motivation Emotion*, vol. 10, pp. 159–168, 1986.
[20] P. Ekman, "Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique," *Psychol. Bull.*, vol. 115, no. 2, pp. 268–287, 1994.
[21] P. Ekman, "Facial expressions," *Handbook Cogn. Emotion*, vol. 16, no. 301, 1999, Art. no. e320.
[22] E. Lee, J. I. Kang, I. H. Park, J.-J. Kim, and S. K. An, "Is a neutral face really evaluated as being emotionally neutral?," *Psychiatry Res.*, vol. 157, no. 1/3, pp. 77–85, 2008.
[23] R. Plutchik, "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *Amer. Scientist*, vol. 89, no. 4, pp. 344–350, 2001.
[24] A. Mehrabian, *Basic Dimensions for a General Psychological Theory: Implications for Personality, Social, Environmental, and Developmental Studies*. Cambridge, U.K.: Oelgeschlager, Gunn & Hain, 1980.
[25] J. A. Russell, "A circumplex model of affect," *J. Pers. Social Psychol.*, vol. 39, no. 6, 1980, Art. no. 1161.
[26] T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," in *Proc. 2nd Int. Conf. Knowl. Capture*, 2003, pp. 70–77.
[27] S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," in *Proc. 20th Int. Conf. Comput. Linguistics*, 2004, pp. 1367–1373.
[28] B. Liu et al., "Sentiment analysis and subjectivity," *Handbook Natural Lang. Process.*, vol. 2, no. 2010, pp. 627–666, 2010.
[29] B. Liu, M. Hu, and J. Cheng, "Opinion observer: Analyzing and comparing opinions on the web," in *Proc. 14th Int. Conf. World Wide Web*, 2005, pp. 342–351.
[30] A. Pak et al., "Twitter as a corpus for sentiment analysis and opinion mining," in *Proc. Lang. Resour. Eval. Conf.*, 2010, pp. 1320–1326.
[31] E. Cambria, D. Das, S. Bandyopadhyay, and A. Feraco, "Affective computing and sentiment analysis," *A practical guide to sentiment Anal.*, pp. 1–10, 2017.
[32] E. Cambria, R. Speer, C. Havasi, and A. Hussain, "Senticnet: A publicly available semantic resource for opinion mining," in *Proc. 2010 AAAI Fall Symp. Ser.*, 2010, pp. 14–18.
[33] H. Salam, O. Celiktutan, H. Gunes, and M. Chetouani, "Automatic context-aware inference of engagement in HMI: A survey," *IEEE Trans. Affect. Comput.*, vol. 15, no. 2, pp. 445–464, Second Quarter, 2024.
[34] R. Arya, J. Singh, and A. Kumar, "A survey of multidisciplinary domains contributing to affective computing," *Comput. Sci. Rev.*, vol. 40, May 2021, doi: 10.1016/j.cosrev.2021.100399.
[35] P. A. Baffour, H. Nunoo-Mensah, E. Keelson, and B. Kommey, "A survey on deep learning algorithms in facial emotion detection and recognition," *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 7, pp. 24–32, 2022.
[36] M. Jampour and M. Javidi, "Multiview facial expression recognition, a survey," *IEEE Trans. Affect. Comput.*, vol. 13, no. 1, pp. 2086–2105, Fourth Quarter, 2022.
[37] S. Li and W. Deng, "Deep facial expression recognition: A survey," *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1195–1215, Third Quarter, 2022.
[38] S. P. Yadav, S. Zaidi, A. Mishra, and V. Yadav, "Survey on machine learning in speech emotion recognition and vision systems using a recurrent neural network (RNN)," *Arch. Comput. Methods Eng.*, vol. 29, pp. 1753–1770, May, 2022.
[39] R. Banse and K. R. Scherer, "Acoustic profiles in vocal emotion expression," *J. Pers. Social Psychol.*, vol. 70, no. 3, pp. 614–636, 1996.
[40] J. Chen et al., "Heu emotion: A large-scale database for multimodal emotion recognition in the wild," *Neural Comput. Appl.*, vol. 33, pp. 8669–8685, 2021.
[41] L. Tabbaa et al., "Vreed: Virtual reality emotion recognition dataset using eye tracking & physiological measures," in *Proc. ACM Interactive Mobile Wearable Ubiquitous Technol.*, vol. 5, no. 4, pp. 1–20, 2021.
[42] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "Affectnet: A database for facial expression, valence, and arousal computing in the wild," *IEEE Trans. Affect. Comput.*, vol. 10, no. 1, pp. 18–31, First Quarter, 2017.
[43] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," *Image Vis. Comput.*, vol. 28, no. 5, pp. 807–813, 2010.
[44] G. Zhao, X. Huang, M. Taini, S. Z. Li, and M. PietikäInen, "Facial expression recognition from near-infrared videos," *Image Vis. Comput.*, vol. 29, no. 9, pp. 607–619, 2011.
[45] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for eeg-based emotion classification," in *Proc. 6th Int. IEEE/EMBS Conf. Neural Eng.*, 2013, pp. 81–84.
[46] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer, "The first facial expression recognition and analysis challenge," in *Proc. 2011 IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2011, pp. 921–926.
[47] T. Bänziger and K. R. Scherer, "Introducing the geneva multimodal emotion portrayal (GEMEP-CS) corpus," *Blueprint Affective Computing: A Sourcebook*, vol. 2010, pp. 271–94, 2010.
[48] S. Haq, P. J. Jackson, and J. Edge, "Audio-visual feature selection and reduction for emotion classification," in *Proc. Int. Conf. Auditory- Vis. Speech Process.*, Tangalooma, Australia, 2008, pp. 185–190.
[49] 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 *Proc. 2010 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.-Workshops*, 2010, pp. 94–101.
[50] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Proc. 4th IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2000, pp. 46–53.
[51] S. Koelstra et al., "DEAP: A database for emotion analysis; using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, First Quarter, 2012.
[52] I. J. Goodfellow et al., "Challenges in representation learning: A report on three machine learning contests," in *Proc. 20th Int. Conf. Neural Inf. Process.*, Daegu, Korea, 2013, pp. 117–124.
[53] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 42–55, First Quarter, 2012.
[54] J. R. Fontaine, K. R. Scherer, E. B. Roesch, and P. C. Ellsworth, "The world of emotions is not two-dimensional," *Psychol. Sci.*, vol. 18, no. 12, pp. 1050–1057, 2007.
[55] M. Valstar et al., "Induced disgust, happiness and surprise: An addition to the MMI facial expression database," in *Proc. 3rd Int. Workshop EMOTION: Corpora for Res. Emotion Affect*, Paris, France, 2010, Art. no. 65.
[56] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expression database for facial behavior research," in 7*Proc. th Int. Conf. Autom. Face Gesture Recognit.*, 2006, pp. 211–216.
[57] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe, "DECAF: MEG-based multimodal database for decoding affective physiological responses," *IEEE Trans. Affect. Comput.*, vol. 6, no. 3, pp. 209–222, Third Quarter 2015.
[58] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *Proc. 3rd IEEE Int. Conf. Autom. Face Gesture Recognit.*, 1998, pp. 200–205.
[59] S. Li, W. Deng, and J. Du, "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2852–2861.

[60] X. Jiang et al., "Dfew: A large-scale database for recognizing dynamic facial expressions in the wild," in *Proc. 28th ACM Int. Conf. Multimedia*, 2020, pp. 2881–2889.
[61] A. Dhall, R. Goecke, S. Ghosh, J. Joshi, J. Hoey, and T. Gedeon, "From individual to group-level emotion recognition: Emotiw 5.0," in *Proc. 19th ACM Int. Conf. Multimodal Interaction*, 2017, pp. 524–528.
[62] J. Machajdik and A. Hanbury, "Affective image classification using features inspired by psychology and art theory," in *Proc. 18th ACM Int. Conf. Multimedia*, 2010, pp. 83–92.
[63] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification," in *Proc. 45th Annu. Meeting Assoc. Comput. linguistics*, 2007, pp. 440–447.
[64] M. Valstar et al., "AVEC 2013: The continuous audio/visual emotion and depression recognition challenge," in *Proc. 3rd ACM Int. Workshop Audio/Visual Emotion Challenge*, 2013, pp. 3–10.
[65] C. Eroglu Erdem, C. Turan, and Z. Aydin, "Baum-2: A multilingual audio-visual affective face database," *Multimedia Tools Appl.*, vol. 74, no. 18, pp. 7429–7459, 2015.
[66] E. Douglas-Cowie, R. Cowie, and M. Schröder, "A new emotion database: Considerations, sources and scope," in *Proc. ISCA Tutorial Res. Workshop Speech Emotion*, 2000, pp. 1–6.
[67] X. Mao, L. Chen, and B. Zhang, "Mandarin speech emotion recognition based on a hybrid of hmm/ann," *Int. J. Comput.*, vol. 1, no. 4, pp. 321–324, 2007.
[68] L. Zhang et al., ""biovid emo db": A multimodal database for emotion analyses validated by subjective ratings," in *Proc. IEEE Symp. Ser. Comput. Intell.*, 2016, pp. 1–6.
[69] P. Melville, W. Gryc, and R. D. Lawrence, "Sentiment analysis of blogs by combining lexical knowledge with text classification," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2009, pp. 1275–1284.
[70] X. Zhang et al., "BP4D-spontaneous: A high-resolution spontaneous 3D dynamic facial expression database," *Image Vis. Comput.*, vol. 32, no. 10, pp. 692–706, 2014.
[71] G. Caridakis, J. Wagner, A. Raouzaiou, Z. Curto, E. Andre, and K. Karpouzis, "A multimodal corpus for gesture expressivity analysis," in *Proc. Multimodal Corpora: Adv. Capturing Coding Analyzing Multimodality*, 2010, pp. 80–85.
[72] H. M. Fayek, M. Lech, and L. Cavedon, "Evaluating deep learning architectures for speech emotion recognition," *Neural Netw.*, vol. 92, pp. 60–68, 2017.
[73] Y. Li, J. Tao, B. Schuller, S. Shan, D. Jiang, and J. Jia, "MEC 2017: Multimodal emotion recognition challenge," in *Proc. 1st Asian Conf. Affect. Comput. Intell. Interaction*, 2018, pp. 1–5.
[74] V. Markova, T. Ganchev, and K. Kalinkov, "Clas: A database for cognitive load, affect and stress recognition," in *Proc. 2019 Int. Conf. Biomed. Innovations Appl.*, 2019, pp. 1–4.
[75] A. Battocchi, F. Pianesi, and D. Goren-Bar, "Dafex: Database of facial expressions," in *Proc. Int. Conf. Intell. Technol. Interactive Entertainment*, 2005, pp. 303–306.
[76] S. Katsigiannis and N. Ramzan, "Dreamer: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 1, pp. 98–107, 2017.
[77] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, Jun. 2005.
[78] A. Vaswani et al., "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 1–11, 2017.
[79] S. Abrilian, L. Devillers, S. Buisine, and J.-C. Martin, "EmoTV1: Annotation of real-life emotions for the specification of multimodal affective interfaces," *HCI Int.*, vol. 401, 2005, pp. 407–408.
[80] H. O'Reilly et al., "The eu-emotion stimulus set: A validation study," *Behav. Res. Methods*, vol. 48, pp. 567–576, 2016.
[81] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 1096–1103.
[82] E. S. Dan-Glauser and K. R. Scherer, "The geneva affective picture database (gaped): A new 730-picture database focusing on valence and normative significance," *Behav. Res. Methods*, vol. 43, pp. 468–477, 2011.
[83] K. R. Scherer and G. Ceschi, "Lost luggage: A field study of emotion–antecedent appraisal," *Motivation Emotion*, vol. 21, pp. 211–235, 1997.
[84] M. Speriosu, N. Sudan, S. Upadhyay, and J. Baldridge, "Twitter polarity classification with label propagation over lexical links and the follower graph," in *Proc. 1st Workshop Unsupervised Learn.*, 2011, pp. 53–63.
[85] L.-P. Morency, R. Mihalcea, and P. Doshi, "Towards multimodal sentiment analysis: Harvesting opinions from the web," in *Proc. 13th Int. Conf. Multimodal Interfaces*, 2011, pp. 169–176.
[86] A. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proc. 49th Annu. Meeting Assoc. Comput. Linguistics: Hum. Lang. Technol.*, 2011, pp. 142–150.
[87] M. Biehl et al., "Matsumoto and ekman's japanese and caucasian facial expressions of emotion (JACFEE): Reliability data and cross-national differences," *J. Nonverbal Behav.*, vol. 21, pp. 3–21, 1997.
[88] C. Y. Park et al., "K-emocon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations," *Sci. Data*, vol. 7, no. 1, 2020, Art. no. 293.
[89] G. Wen et al., "Random deep belief networks for recognizing emotions from speech signals," *Comput. Intell. Neurosci.*, vol. 2017, pp. 1–9, 2017.
[90] K. Akuzawa, Y. Iwasawa, and Y. Matsuo, "Expressive speech synthesis via modeling expressions with variational autoencoder," 2019, *arXiv: 1804.02135*.
[91] S. Poria, D. Hazarika, N. Majumder, G. Naik, E. Cambria, and R. Mihalcea, "Meld: A multimodal multi-party dataset for emotion recognition in conversations," 2018, *arXiv: 1810.02508*.
[92] G. Shen, X. Wang, X. Duan, H. Li, and W. Zhu, "Memor: A dataset for multimodal emotion reasoning in videos," in *Proc. 28th ACM Int. Conf. Multimedia*, 2020, pp. 493–502.
[93] Z. Zhang et al., "Multimodal spontaneous emotion corpus for human behavior analysis," in *Proc. IEEE Conf. comput. Vis. Pattern Recognit.*, 2016, pp. 3438–3446.
[94] V. Pérez-Rosas, R. Mihalcea, and L.-P. Morency, "Utterance-level multimodal sentiment analysis," in *Proc. 51st Annu. Meeting Assoc. Comput. Linguistics*, 2013, pp. 973–982.
[95] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," 2002, *arXiv:cs/0205070*.
[96] T. Song, W. Zheng, C. Lu, Y. Zong, X. Zhang, and Z. Cui, "Mped: A multimodal physiological emotion database for discrete emotion recognition," *IEEE Access*, vol. 7, pp. 12177–12191, 2019.
[97] S. Castro, D. Hazarika, V. Pérez-Rosas, R. Zimmermann, R. Mihalcea, and S. Poria, "Towards multimodal sarcasm detection (an_obviously_perfect paper)," 2019, *arXiv: 1906.01815*.
[98] H.-C. Chou, W.-C. Lin, L.-C. Chang, C.-C. Li, H.-P. Ma, and C.-C. Lee, "Nnime: The nthu-ntua chinese interactive multimodal emotion corpus," in *Proc. 7th Int. Conf. Affect. Comput. Intell. Interaction*, 2017, pp. 292–298.
[99] D. A. Shamma, L. Kennedy, and E. F. Churchill, "Tweet the debates: Understanding community annotation of uncollected sources," in *Proc. 1st SIGMM Workshop Social Media*, 2009, pp. 3–10.
[100] K. Zhang, H. Zhang, S. Li, C. Yang, and L. Sun, "The PMEmo dataset for music emotion recognition," in *Proc. 2018 ACM Int. Conf. Multimedia Retrieval*, 2018, pp. 135–142.
[101] O. Perepelkina, E. Kazimirova, and M. Konstantinova, "Ramas: Russian multimodal corpus of dyadic interaction for affective computing," in *Proc. Int. Conf. Speech Comput.*, 2018, pp. 501–510.
[102] S. R. Livingstone and F. A. Russo, "The ryerson audio-visual database of emotional speech and song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in north american english," *PLoS One*, vol. 13, no. 5, 2018, Art. no. e0196391.
[103] P. Greasley, C. Sherrard, and M. Waterman, "Emotion in language and speech: Methodological issues in naturalistic approaches," *Lang. Speech*, vol. 43, no. 4, pp. 355–375, 2000.
[104] Y. Wang and L. Guan, "Recognizing human emotional state from audiovisual signals," *IEEE Trans. Multimedia*, vol. 10, no. 5, pp. 936–946, Aug. 2008.
[105] H. Gunes and M. Pantic, "Dimensional emotion prediction from spontaneous head gestures for interaction with sensitive artificial listeners," in *Proc. 10th Int. Con. Intell. Virtual Agents*, Philadelphia, PA, USA, 2010, pp. 371–377.
[106] A. K. Davison, C. Lansley, N. Costen, K. Tan, and M. H. Yap, "SAMM: A spontaneous micro-facial movement dataset," *IEEE Trans. Affect. Comput.*, vol. 9, no. 1, pp. 116–129, First Quarter, 2018.
[107] W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, and A. Cichocki, "Emotionmeter: A multimodal framework for recognizing human emotions," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 1110–1122, 2018.
[108] T.-H. Li, W. Liu, W.-L. Zheng, and B.-L. Lu, "Classification of five emotions from EEG and eye movement signals: Discrimination ability and stability over time," in *Proc. 9th Int. IEEE/EMBS Conf. Neural Eng.*, 2019, pp. 607–610.
[109] W.-L. Zheng and B.-L. Lu, "A multimodal approach to estimating vigilance using eeg and forehead EOG," *J. Neural Eng.*, vol. 14, no. 2, 2017, Art. no. 026017.

[110] G. McKeown, M. F. Valstar, R. Cowie, and M. Pantic, “The semaine corpus of emotionally coloured character interactions,” in *Proc. 2010 IEEE Int. Conf. Multimedia Expo*, 2010, pp. 1079–1084.

[111] T. Ganchev, V. Markova, I. Lefterov, and Y. Kalinin, “Overall design of the slade data acquisition system,” in *Proc. 2nd Int. Sci. Conf. Intell. Inf. Technol. Ind.*, 2018, pp. 56–65.

[112] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikäinen, “A spontaneous micro-expression database: Inducement, collection and baseline,” in *Proc. 10th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit.*, 2013, pp. 1–6.

[113] A. Go, R. Bhayani, and L. Huang, “Twitter sentiment classification using distant supervision,” *CS224N Project Rep.*, Stanford, vol. 1, no. 12, 2009, Art. no. 200.9.

[114] G. E. Hinton, “Learning multiple layers of representation,” *Trends Cogn. Sci.*, vol. 11, no. 10, pp. 428–434, 2007.

[115] B. Schuller et al., “Being bored? Recognising natural interest by extensive audiovisual integration for real-life application,” *Image Vis. Comput.*, vol. 27, no. 12, pp. 1760–1774, 2009.

[116] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, “Improved training of Wasserstein GANs,” in *Proc. Adv. Neural Inf. Process. Syst.*, I. Guyon et al., Eds. 2017, vol. 30, pp. 1–11.

[117] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, “Introducing wesad, a multimodal dataset for wearable stress and affect detection,” in *Proc. 20th ACM Int. Conf. Multimodal Interaction*, 2018, pp. 400–408.

[118] R. Mihalcea, “Towards multimodal sentiment analysis,” 2011.

[119] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, “AMIGOS: A dataset for affect, personality and mood research on individuals and groups,” *IEEE Trans. Affect. Comput.*, vol. 12, no. 2, pp. 479–493, Second Quarter, 2018.

[120] H. Gunes and M. Piccardi, “A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior,” in *Proc. 18th Int. Conf. Pattern Recognit.*, 2006, pp. 1148–1153.

[121] E. Douglas-Cowie et al., “The humaine database: Addressing the collection and annotation of naturalistic and induced emotional data,” in *Proc. 2nd Int. Conf. Affect. Comput. Intell. Interaction*, Lisbon, Portugal, 2007, pp. 488–500.

[122] O. Langner, R. Dotsch, G. Bijlstra, D. H. Wigboldus, S. T. Hawk, and A. Van Knippenberg, “Presentation and validation of the radboud faces database,” *Cogn. Emotion*, vol. 24, no. 8, pp. 1377–1388, 2010.

[123] F. Ringeval, A. Sonderegger, J. Sauer, and D. Lalanne, “Introducing the recola multimodal corpus of remote collaborative and affective interactions,” in *Proc. 10th IEEE Int. Conf. workshops Autom. Face Gesture Recognit.*, 2013, pp. 1–8.

[124] A. Dhall, R. Goecke, S. Lucey, and T. Gedeon, “Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark,” in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2011, pp. 2106–2112.

[125] S. Cheng, I. Kotsia, M. Pantic, and S. Zafeiriou, “4DFAB: A large scale 4D database for facial expression analysis and biometric applications,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 5117–5126.

[126] A. Dhall et al., “Collecting large, richly annotated facial-expression databases from movies,” *IEEE Multimedia*, vol. 19, no. 3, Sep. 2012, Art. no. 34.

[127] R. Subramanian, J. Wache, M. K. Abadi, R. L. Vieriu, S. Winkler, and N. Sebe, “ASCERTAIN: Emotion and personality recognition using commercial sensors,” *IEEE Trans. Affect. Comput.*, vol. 9, no. 2, pp. 147–160, Second Quarter, 2016.

[128] M. Valstar et al., “Avec 2014: 3D dimensional affect and depression recognition challenge,” in *Proc. 4th Int. Workshop Audio/Visual Emotion Challenge*, 2014, pp. 3–10.

[129] S. Zhalehpour, O. Onder, Z. Akhtar, and C. E. Erdem, “BAUM-1: A spontaneous audio-visual face database of affective and mental states,” *IEEE Trans. Affect. Comput.*, vol. 8, no. 3, pp. 300–313, Third Quarter, 2017.

[130] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale, “A high-resolution 3d dynamic facial expression database,” in *Proc. 8th IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2008, pp. 1–6.

[131] W.-J. Yan et al., “Casme ii: An improved spontaneous micro-expression database and the baseline evaluation,” *PLoS One*, vol. 9, no. 1, 2014, Art. no. e86041.

[132] Y. Li, J. Tao, L. Chao, W. Bao, and Y. Liu, “Cheavd: A chinese natural emotional audio–visual database,” *J. Ambient Intell. Humanized Comput.*, vol. 8, pp. 913–924, 2017.

[133] A. B. Zadeh, P. P. Liang, S. Poria, E. Cambria, and L.-P. Morency, “Multimodal language analysis in the wild: Cmu-mosei dataset and interpretable dynamic fusion graph,” in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, 2018, pp. 2236–2246.

[134] A. Metallinou et al., “The usc creativeit database: A multimodal database of theatrical improvisation,” *Multimodal Corpora: Adv. Capturing Coding Analyzing Multimodality*, 2010, Art. no. 55.

[135] I. S. Engberg, A. V. Hansen, O. Andersen, and P. Dalsgaard, “Design, recording and verification of a danish emotional speech database,” in *Proc. 5th Eur. Conf. Speech Commun. Technol.*, 1997.

[136] N. Fourati and C. Pelachaud, “Emilya: Emotional body expression in daily actions database,” in *Proc. Lang. Resour. Eval. Conf.*, 2014, pp. 3486–3493.

[137] F. Burkhardt et al., “A database of german emotional speech,” in *Proc. Interspeech*, 2005, pp. 1517–1520.

[138] H. Ranganathan, S. Chakraborty, and S. Panchanathan, “Multimodal emotion recognition using deep learning architectures,” in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2016, pp. 1–9.

[139] C. F. Benitez-Quiroz, R. Srinivasan, and A. M. Martinez, “Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 5562–5570.

[140] O. Martin, I. Kotsia, B. Macq, and I. Pitas, “The enterface05 audio-visual emotion database,” in *Proc. 22nd Int. Conf. Data Eng. Workshops*, 2006, pp. 8–8.

[141] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, “From facial expression recognition to interpersonal relation prediction,” *Int. J. Comput. Vis.*, vol. 126, pp. 550–569, 2018.

[142] M. Wöllmer et al., “Youtube movie reviews: Sentiment analysis in an audio-visual context,” *IEEE Intell. Syst.*, vol. 28, no. 3, pp. 46–53, May/Jun. 2013.

[143] C. Busso et al., “Iemocap: Interactive emotional dyadic motion capture database,” *Lang. Resour. Eval.*, vol. 42, pp. 335–359, 2008.

[144] E. Goeleven, R. De Raedt, L. Leyman, and B. Verschuere, “The karolinska directed emotional faces: A validation study,” *Cogn. Emotion*, vol. 22, no. 6, pp. 1094–1118, 2008.

[145] Y. Baveye, E. Dellandrea, C. Chamaret, and L. Chen, “LIRIS-ACCEDE: A video database for affective content analysis,” *IEEE Trans. Affect. Comput.*, vol. 6, no. 1, pp. 43–55, First Quarter, 2015.

[146] A. Dhall, O. Ramana Murthy, R. Goecke, J. Joshi, and T. Gedeon, “Video and image based emotion recognition challenges in the wild: Emotiw 2015,” in *Proc. ACM Int. Conf. Multimodal Interaction*, 2015, pp. 423–426.

[147] M. Kipp and J.-C. Martin, “Gesture and emotion: Can basic gestural form features discriminate emotions?,” in *Proc. 3rd Int. Conf. Affect. Comput. Intell. Interact. Workshops*, 2009, pp. 1–8.

[148] M. Grimm, K. Kroschel, and S. Narayanan, “The vera am mittag german audio-visual emotional speech database,” in *Proc. 2008 IEEE Int. Conf. Multimedia Expo*, 2008, pp. 865–868.

Pedro J. Vaz received the BSc degree in electrical and electronics engineering and the MSc degree in electrical and computer engineering from the University of Algarve, Faro, Portugal, in 2009 and 20022, respectively. He is currently working toward the PhD degree in informatics engineering with the University of Algarve. From 2009 to 2010 he worked as a software engineer with Maine Avenue Technologies, in Madrid, Spain. From 2011 to 2015 he worked as an electrical engineer, and from 2015 up to 2016 as a senior electronics design engineer with Stanley Black & Decker GmbH, in Idstein, Germany. From 2018 to 2021 he was a high school teacher, lecturing classes in Computer Science, Electricity and Electronics fields. From 2021 to 2023 he worked as a researcher under the project Green Spaces SMART Irrigation Control (GSSIC), and since 2023 he has been a doctoral scholarship holder, under the Sustainable Horizons European Universities (SHEs) program. Since 2021 he has been an Invited Assistant, lecturing Computer Programming classes with the University of Algarve. Starting in January 2024 he joined the Research Centre NOVA LINCS - Algarve. His research interests include machine learning, affective computing, data science, electronics, and audio signal processing. Details at https://tinyurl.com/554krh5n

João M. F. Rodrigues received the bachelor's degree in electrical engineering in 1993, the master's degree in systems and computer engineering in 1998, the PhD degree in electronic and computer engineering with a speciality in computer science, in 2008, and the habilitation degree in electrical and computer engineering. He is currently serving as the vice-eector for Transfer, Innovation, and Digital University at the University of the Algarve (UAlg) in Portugal. At present is a coordinator professor in tenure with the Institute of Engineering (ISE), UAlg. Since 1994 he has taught Curricular Units in Computer Science and Computer Vision. In addition to his academic and administrative roles, Professor Rodrigues is a member of the Research Centre NOVA LINCS - Algarve. He has extensive research experience, having participated in more than 30 nationally or internationally funded scientific projects, several of which he served as coordinator. He has co-author more than 200 scientific papers and is a member of the editorial board of several international journals. His main areas of interest include computer vision, human-computer interaction, and human-centered Artificial Intelligence and Affective Computing, focusing on adaptive interfaces, emotion recognition, human-sense modelling, and human-machine cooperation. Details at https://tinyurl.com/56fvsuwb.

Pedro J. S. Cardoso received the BSc degree in mathematics/computer science from the University of Coimbra, Portugal, the MSc degree in computational mathematics from the University of Minho, Portugal, and the PhD degree in mathematics/operations research from the University of Seville, Spain. Serving as a coordinator professor and researcher with the Universidade do Algarve for over two decades, he is also an integrated member of the NOVA LINCS research center. His extensive expertise in computer science has led him to specialize in machine learning, data science, and multiple objective meta-heuristics. He has a long-standing career as a lecturer in these fields. Throughout his career, he has been involved in over a dozen scientific and development projects and has published more than ninety papers. Additionally, he has edited over a dozen books. His primary research interests include machine learning, particularly its real-world applications, affective computing, anomaly detection, and energy efficiency. Details at https://tinyurl.com/26hdrntd.