

Automated Affect and Emotion Recognition from Cardiovascular Signals – A Systematic Overview Of The Field

Paweł Jemioło

AGH University of Science and Technology
al. A. Mickiewicza 30, 30-059 Krakow, Poland
pawljmlo@agh.edu.pl

Dawid Storman

Jagiellonian University Medical College,
ul. Św. Anny 12, 31-008 Krakow, Poland
dawid.storman@doctoral.uj.edu.pl

Maria Mamica and Mateusz Szymkowski
AGH University of Science and Technology
al. A. Mickiewicza 30, 30-059 Krakow, Poland
{mamica,szymkows}@student.agh.edu.pl

Patryk Orzechowski
University of Pennsylvania
3700 Hamilton Walk, Philadelphia 19104, USA
AGH University of Science and Technology
al. A. Mickiewicza 30, 30-059 Krakow, Poland
patryk.orzechowski@gmail.com

Abstract

Currently, artificial intelligence is increasingly used to recognize and differentiate emotions. Through the action of the nervous system, the heart and vascular system can respond differently depending on the type of arousal. With the growing popularity of wearable devices able to measure such signals, people may monitor their states and manage their wellness. Our goal was to explore and summarize the field of automated emotion and affect recognition from cardiovascular signals.

According to our protocol, we searched electronic sources (MEDLINE, EMBASE, Web of Science, Scopus, dblp, Cochrane Library, IEEE Explore, arXiv and medRxiv) up to 31 August 2020. In the case of all identified studies, two independent reviewers were involved at each stage: screening, full-text assessment, data extraction, and quality evaluation. All conflicts were resolved during the discussion. The credibility of included studies was evaluated using a proprietary tool based on QUADAS, PROBAST. After screening 4649 references, we identified 195 eligible studies.

From artificial intelligence most used methods in emotion or affect recognition were Support Vector Machines (42.86%), Neural Network (21.43%), and k-Nearest Neighbors (11.67%). Among the most explored datasets were DEAP (10.26%), MAHNOB-HCI (10.26%), AMIGOS (6.67%) and DREAMER (2.56%). The most frequent cardiovascular signals involved electrocardiogram (63.16%), photoplethysmogram (15.79%), blood volume pressure (13.16%) and heart rate (6.58%).

Sadness, fear, and anger were the most examined emotions. However, there is no standard set of investigated internal feelings. On average, authors explore 4.50 states (range from 4 to 24 feelings).

Research using artificial intelligence in recognizing emotions or affect using cardiovascular signals shows an upward trend. There are significant variations in the quality of the datasets, the choice of states to detect, and the classifiers used for analysis.

Research project supported by program Excellence initiative – research university for the University of Science and Technology. The authors declare that they have no conflict of interest.

1. Introduction

Emotions play an essential role in many mental processes, such as decision-making, perception or learning. They impact both the physiology and behaviour of a person [1].

Since the end of the 19th century, multiple theories on emotional expressions have been formulated [2]. As many believe, emotions are inherent and biologically connected from the moment of birth and that they are universal [3]. Ekman's theory is often cited in emotion recognition as it is well suited to mapping the various reactions of the body to mental states. At the opposite extreme, there is the theory of the core affect proposed by Russell [4]. In that understanding, emotions are described via valence, arousal and sometimes dominance [4].

Barrett expanded the understanding of emotions by describing that emotions are constructed from past experiences and socialization [5]. In that sense, emotions cannot be considered universal, and what seems natural for people is mentioned earlier affect [5].

Wierzbicka [6] made another critical point in the debate on emotions. As she claims, new terms need to be established as all words we use to describe our states are culturally loaded. What she proposes is a metalanguage based on universal primitives [6].

The diversity of theories has been followed by the increased number of areas in which emotion or affect recognition is taken advantage of, including interaction with people and robotics [7–9], intelligent vehicles [10, 11], psychiatry and psychotherapy [12, 13], and many others. Machine learning (ML) can be utilized for patients experiencing alexithymia accompanying other disorders (autism spectrum disorders, somatoform disorders), i.e. inability to become aware, understand and express their own emotions [14]. Systems enabling automated emotion recognition may be beneficial in everyday social interactions and during therapy, serving as a learning tool for associating emotions [15].

Due to a sharp increase in the development of wearable equipment, automated emotion recognition might also be applied for monitoring emotional status constantly. Such technology is known for improving disease management and allowing a better understanding of the patients' needs [16]. Similarly, assisted living and wellness management are at one's fingertips with emotion or affect recognition incorporated in smartphone applications.

What is more, automated emotion recognition may also be applied in intelligent assistants, enabling more natural conversations [17, 18]. It can also be used in the entertainment industry [19, 20], enhancing the level of immersion, e.g. in games [21].

Recognition of human internal states has been performed using different facial expressions [22], speech [23], EEG [24], cardiovascular [25], or other physiological signals [26].

Over the years, artificial intelligence (AI) techniques and ML were applied to emotions classification. The most popular ML approaches applied to the discussed field include Support Vector Machine (SVM) [27, 28], k-nearest neighbours (kNN) [29], and, more recently, deep neural networks [22, 30], which allow for more flexibility and reduce the time needed for data preparation. However, the popularity of specific approaches does not necessarily translate to the actual effectiveness of the methods in identifying emotions or affect. In order to make such a claim, the methods would need to be compared across multiple datasets, similarly

to OpenML [31] or PMLB [32]. Objective comparison of ML methods would optimally require testing them using the same workflow.

A separate issue is the reproducibility of the analyses under different platforms [33, 34]. The vast majority of the studies neither publish the datasets, nor the source code, which poses challenges in benchmarking. What is more, many authors do not provide all sufficient information on the proposed methods [35]. Unfortunately, we are witnessing the replication crisis not only in computer science but also in psychology. Many studies, previously considered credible, have been criticized on account of methodological flaws [36, 37].

In this paper, we summarize the recent advances in emotion and affect recognition, focusing on models that utilize data from cardiovascular signals (CVSs).

There is a couple of reasons for narrowing the scope of this review. We believe that a targeted study will be more concise when providing implications for practice and future works through homogeneous synthesis. What is more, wearable devices capable of measuring Heart Rate (HR) or even performing Electrocardiography (ECG) are becoming more prevalent [38]. It opens new possibilities for detecting how a given person feels at a particular time and potentially offering relaxation techniques in case of undesirable conditions enabling management of wellness.

2. Research Questions

In this article, we have detailed the predefined research questions [39] as follows:

1. What are the AI methods used for automated emotion or affect recognition from CVSs?
2. What kind of CVSs are used for automated emotion or affect recognition?
3. What are the available datasets used for automated emotion or affect recognition from CVSs?
4. What kind of measuring devices are used for automated emotion or affect recognition from CVSs?
5. What kind of automated emotions or affects are recognized from CVSs?
6. What are the bibliometric data of studies published in automated emotion or affect recognition from CVSs?

3. Methods

3.1. Pre-Search, Protocol, Eligibility Criteria

Pre-search was conducted in August 2020. Our goal was to see if any articles would pass our criteria (see below) and be considered eligible for this review. There were several papers that could be easily traced via Google Scholar.

We published the protocol (see Appendix 1) of the current study on the Open Science Framework (OSF) [39] and then registered it there [40]. The publication and registration were done on 18 March 2021.

The eligible criteria were as follows. We included any paper utilizing AI methods and CVSs for emotion or affect recognition in a laboratory setting. We focused on the performance of such systems, e.g. accuracy, sensitivity, specificity. The population had to include less than half of the participants with a specific feature (e.g., participants with illness, children) for inclusion. We excluded reviews, case studies, introduction to a section or special issue in a journal, and letters to editors. Furthermore, we excluded references contacting entire proceeding and post-conference books. However, we did not exclude particular chapters as they were still present in our search (overlapping).

3.2. Search methods

Both free-text words and MeSH terms were included in the search strategy. We fused terms regarding AI, emotion recognition, and CVSs. This primary search strategy (see Appendix 1, OSF [40]) was developed for MEDLINE and then adapted for other sources searched.

MEDLINE, EMBASE, Web of Science, Scopus, dblp, Cochrane Library, IEEE Explore, arXiv and medRxiv were searched since inception to 31 August 2020. Language and date restrictions were not imposed.

Moreover, we searched the references of included papers in order to identify eligible studies. Such records were then rigorously screened.

3.3. Definitions

We adapted the definition of emotion or affect recognition provided by Kim et al. [41] and Konar et al. [42], which refers to utilizing specific signals (e.g. physiological, behavioural) to find patterns corresponding to specific emotions or affects. We considered papers to be eligible if they included CVSs (separated or in combination with others) as defined by Shu et al. [26, 43]: arterial blood pressure (ABP), pulse oximetry (POX), intracranial pressure (ICP),

heart rate (HR), heart rate variability (HRV), pulse pressure variation (PPV), blood volume pressure (BVP), electrocardiogram (ECG), and photoplethysmogram (PPG). By wearables, we understood devices that use the human body as support (e.g. clothes, accessories), which are independent and comfortable [44].

The following definitions were applied to the computer science domain. Artificial Intelligence, which refers to *the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings* (in this case, emotion recognition) [45]. Deep Learning (DL), which refers to Neural Networks (NN) architectures with more than one hidden layer [46]. Performance metrics, which refer to evaluation methods and frameworks used to compare predictions with actual data [47].

3.4. Data collection

The process of deduplication was conducted using Endnote (Clarivate Analytics®) and Rayyan [48] software. Using the latter application, remaining records were then separately screened by PJ, DS, MM, MS and PO (in pairs). Finally, PJ, DS, MM, MS and PO checked the full texts independently. At this stage, we examined if the papers meet our eligibility criteria.

Then, we specified an extraction form in order to gather all necessary data. Following the protocol (Appendix 1, OSF [40]), we mainly looked into characteristics of included studies – emotions or affect recognized and utilized AI methods, as well as additional analyses, e.g. interpretability. Additionally, we focused on bibliometric data (authors, year and type of publication). We also extracted information about the protocol, code availability, funding and conflict of interest reporting.

PJ, DS, MM, MS, PO and WD gathered all data using mentioned form independently. If the data had been unclear or missing, we planned to contact with study author for clarification.

Each of the specified above stages (titles and abstract screening, full texts assessment and data gathering process) was conducted after pilot practice. If the variance occurred, we discussed it and achieved consensus.

3.5. Quality assessment

We assessed the methodological credibility using the author's tool (see Appendix 4), which was adopted from QUADAS-2 [49], PROBAST [50], and the method proposed by Benton et al. [51]. The evaluation was conducted independently by PJ, DS, MM and MS. All discrepancies were resolved by discussion.

3.6. Analyses

In this review, we focus on qualitative synthesis of included papers regarding utilized signals, measuring devices, recognized emotions and affect, AI models, and bibliometric data.

Further quantitative and additional analyses regarding publicly available datasets will be published separately. We, thus, do not provide an investigation of heterogeneity, sensitivity, nor subgroup analyses. In the upcoming review, we are going to present a complete credibility evaluation.

4. Results

Searches in electronic databases yielded 4649 references (after deduplication), which were then screened by title and abstract. We also hand-searched references using the snowball method and Google Scholar, which resulted in 37 additional references. Next, we assessed 244 full texts for eligibility. Finally, we included 195 papers (see Appendix 2, OSF [39]) for qualitative synthesis. A list of excluded studies and reasons can be found in Appendix 3, OSF [39]. Our review is consistent with PRISMA DTA guidelines [52]. The full study flow described above is presented in Figure 1.

The authors of the studies were affiliated with China (9.92%), South Korea (7.44%) and USA (7.44%) institutions. 41.54% of the papers were published in scientific journals (with no Impact Factor (IF) or with IF ranging from 0.806 to 17.86, mean = 3.55), while the rest was presented at conferences. The oldest study included was published in 2004 and the newest in 2020. In Figure 2, we present how many studies were issued in a particular year.

The most popular journal was *IEEE Transactions on Affective Computing* (17.28%), and the most general conference was the *Conference of the IEEE Engineering in Medicine and Biology Society* (17.46%). 63.08% of included papers reported on funding, 17.44% reported on conflict of interest. The protocol of the studies was registered by authors of 3 (1.54%) studies. In 4 articles, the authors provided the source code of their experiments. We received source code from some authors of included publications after e-mail contact.

93.5% of included papers explored only classification techniques, 5.13% focused primarily on regression models, and the rest conducted both analyses. The most utilized technique was SVM (42.86% of all models used), NN (21.43%), and kNN (11.67%). 18.4% of studies utilized deep techniques, while in the case of 8.96%, it was not clear.

Interestingly, study authors reported using deep methods only in terms of 13.57% of all papers included in this review.

33.85% of studies used only CVSSs for recognition. 42.56% fused information from CVSSs with other physiological or behavioural signals, while the rest of the papers presented both approaches. Authors carried out only their own experiments in 73.85% of included studies. 3.08% used own and publicly available data, while the rest utilized solely the latter. The most explored datasets were DEAP [53], MAHNOB-HCI [54], AMIGOS [55] and DREAMER [56]. They were explored by 10.26%, 10.26%, 6.67%, and 2.56% of all included studies, respectively. The authors used multiple (more than one) databases in 14 primary studies (7.17%).

Among authors' own experiments, the most popular were BIOPAC devices (especially MP100, MP150), which were utilized in 28.21% of included papers. On the other hand, the authors of 11.28% of studies did not report on the name of the used instrument. 7.69% of studies used wearable devices. However, authors reported using such devices in 13.85% of works. Authors utilized mainly the following CVSSs: ECG (63.16%), PPG (15.79%), BVP (13.16%) and HR (6.58%).

Finally, we present results regarding recognized emotions and affect. In Figure 3, we visualized discrete states that were analyzed in included papers most often. The emotion recognition was conducted in 47.18% of papers, while affect was recognized in 48.72% of articles. The remaining studies considered both emotion and affect recognition. However, the actual experiment does not converge with the authors' declaration from the title as they reported conducting affect and emotion recognition in 77.95% and 20.51% studies, respectively. The remaining papers claimed to focus on other states. When it comes to affect, the most explored dimension was valence (38.97% of included papers, with 1-10 classes, mean = 2.47) and arousal (33.82%, with 1-10 classes, mean = 2.52).

In general, the authors recognized 4.50 states on average (range: 2-24). The most typical explored set of emotions included sadness, fear, happiness, neutral, and anger.

The complete results regarding population, outcomes – AI methods performance, additional analyses and quality will be available in a comprehensive review. In a separate paper, we cover the datasets with cardiovascular signals, assess their quality, and summarize the field descriptively.

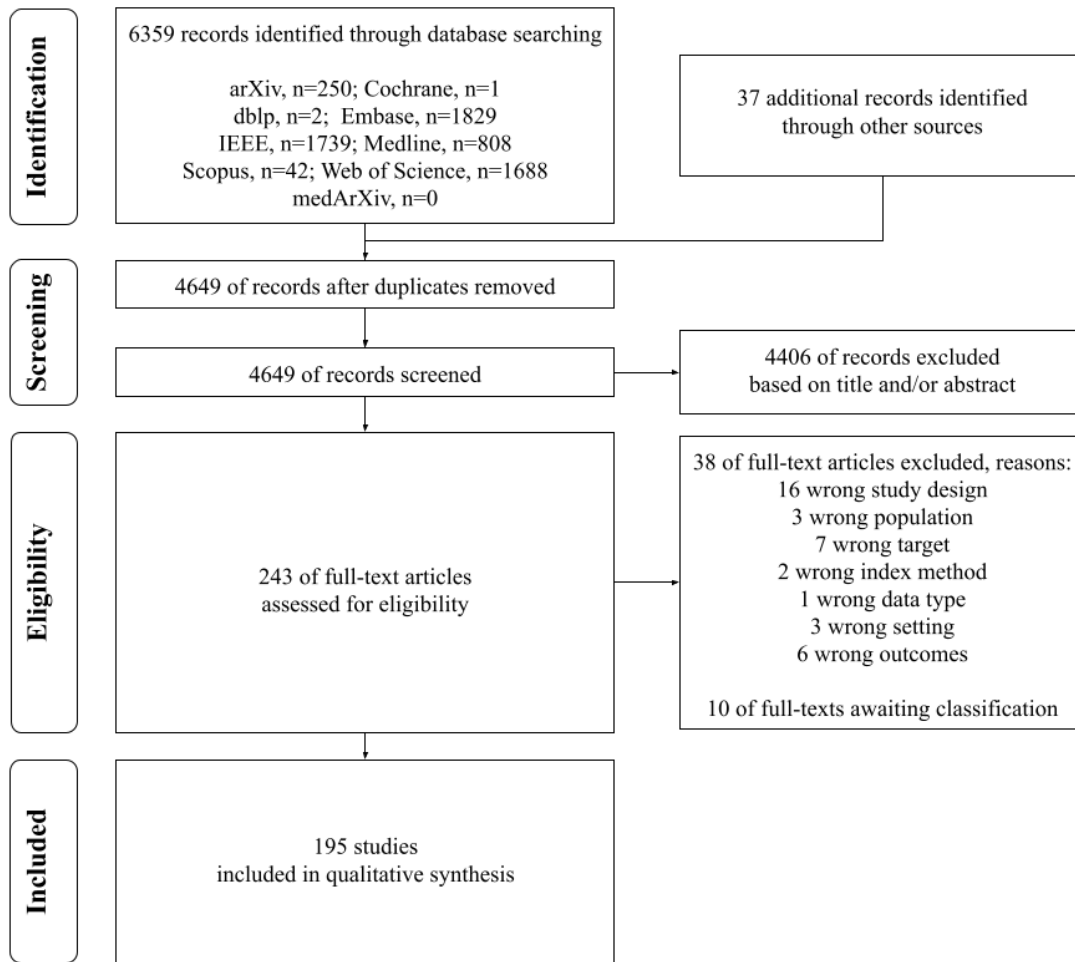


Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) study flow diagram.

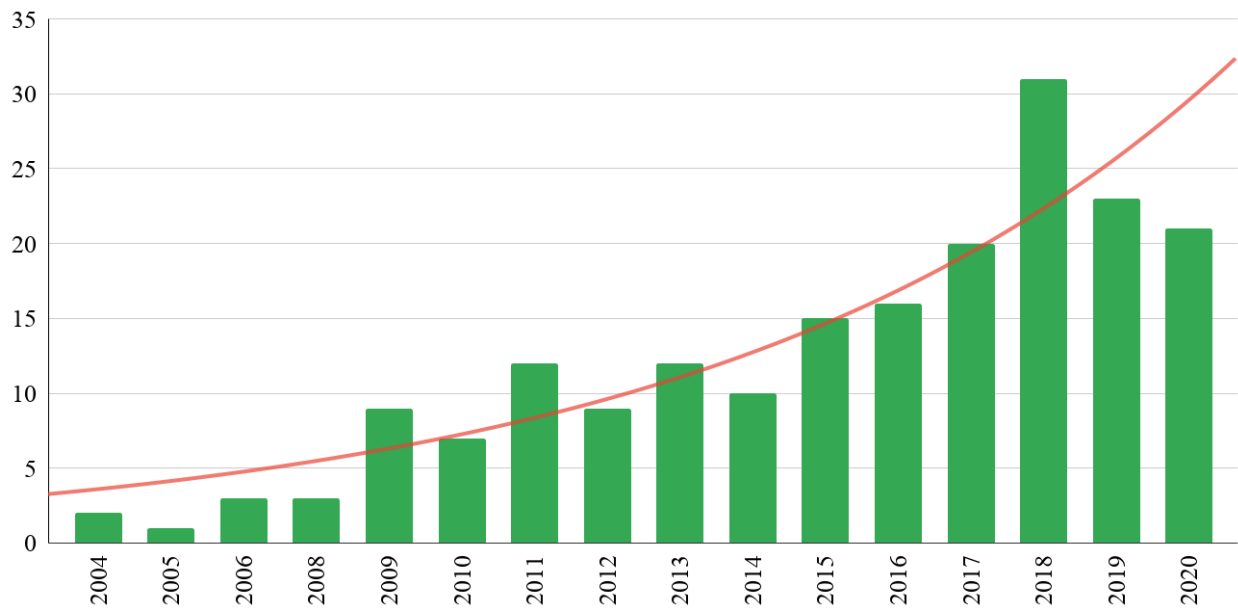


Figure 2. Dates of publication of included studies with trend line.

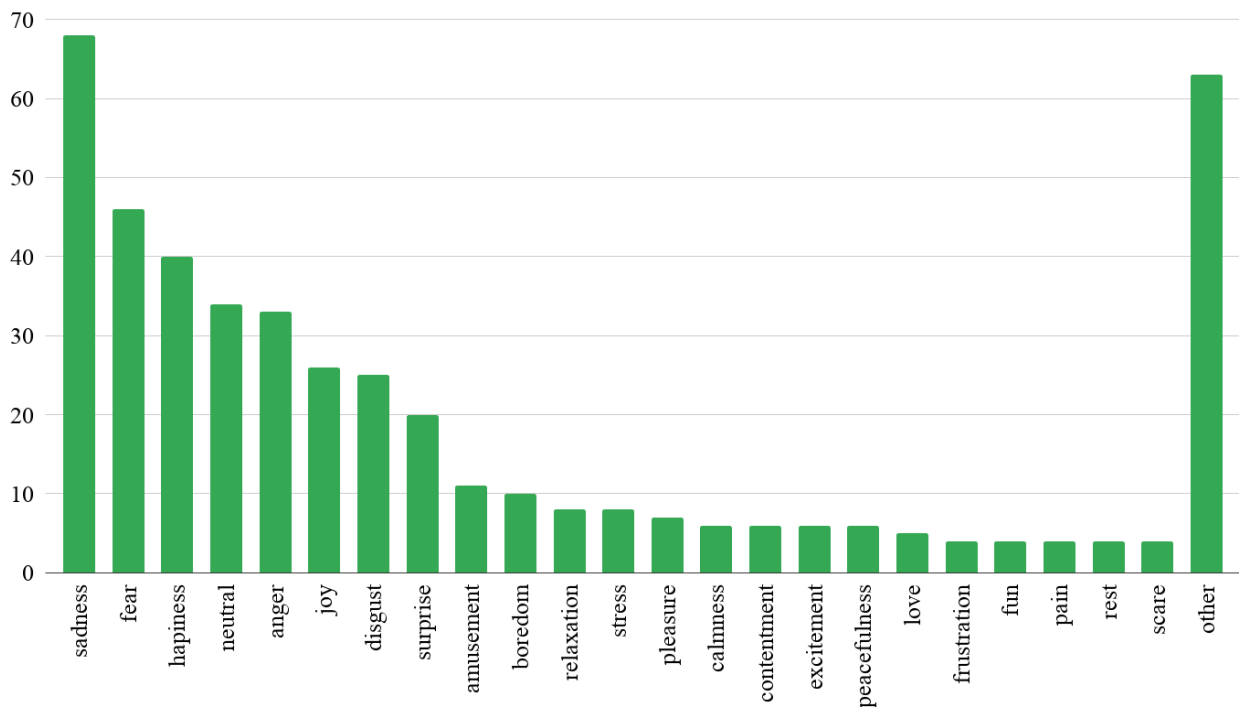


Figure 3. Most frequent states recognized among recognition experiments.

5. Discussion

In the included studies, similarly to Egger et al. [57], we have observed great diversity in the quality of the datasets, choice of emotions to be detected, and classifiers used for analysis. Our study shows a substantial need for improving good scientific practices in this widely explored field of science.

Although the number of publicly available datasets is systematically growing [58], within 18 works, we have identified 18 datasets shared by the authors, which were later reused in 45 works. What is more, the authors compared their elicitation experiments to public data merely in six studies. That impedes direct comparisons between the utilized techniques.

Lack of availability of the data or source code slows down further comparative analysis on the data, which in turn impedes the development and testing of new methods. It also justifies raising questions about the methodological validity of the studies, reproducibility and accuracy of conducted analyses [36]. Moreover, novel algorithms are not commonly tested on more than a single (optimally several) dataset to empirically confirm the outcomes. Thus, some of the claims of their performance are probably overhyped. Despite creating many models and methodologies on the collected datasets, applying these AI architectures in real-life scenarios is still a challenge.

Even though wearable devices are becoming more precise, the quality of medical-level equipment is still superior [59]. What poses a greater concern is that researchers focus on the latter, making adopting automated emotion recognition systems to applications we use every day more complicated. It remains one of the challenges standing in the way of progressing.

Another bothering aspect of the research is the lack of information on potential conflicts of interest or funding sources. What is more, in many primary studies, the authors did not register the protocol of the human subjects experiment before collecting the data. It opens further discussion on the data quality – if the data were collected rigorously, appropriate confounders were controlled or the authors utilized well-established emotion and affect theories [5]. All those factors combined open questions on the trustworthiness of the research and the findings [36]. To mitigate this issue, we should significantly improve reporting in research without delay.

Discrepancies between declared affect or emotion recognition indicate that these two concepts are often confused. They are crucial for understanding the nature of the internal states of a person [5] and should be defined by authors. We agree with Wierzbicka [6] that there is a need for neutral metalanguage in the research of emotions, and our results seem to account for it. We thus urge the authors for coming up with one.

The choice of ML methods may also not be optimal. For example, the vast majority, especially earlier works, take advantage of SVM or their variants to classify emotions and affect. While this classifier is usually sufficient for the authors' needs, it is also known for excessive running times, non-resistance to noise, and general inapplicability to big data [60]. While authors still utilize simple ML algorithms, more and more focus is put on DL methods, which proved to be effective in emotion and affect recognition [61,62].

Despite the lack of justification behind the chosen ML or DL model other than an arbitrarily selected metric, researchers usually do not provide any explainability or interpretability in their works. If affective computing is to gain practical use at some point, it is expected that researchers will abandon the black-box approach in human and AI interaction.

As affective computing spans computer science and psychology, researchers require fundamental knowledge in both of these fields. Taking into consideration their rapid development, following the newest trends and scientific achievements may be highly difficult. It appears to be a big challenge for progressing in automated emotion recognition, as many methodological and technological flaws may be observed in publications.

Presented work provides many benefits for anyone researching affective computing as well as adapting existing solutions. The review points out not only unexplored research areas but also recurring methodological and technological flaws. What is more, it lists the AI models solving the proposed affective computing problem and provides the reader with their results. Such information may be highly beneficial for anyone willing to adopt existing algorithms in automated emotion or affect recognition systems.

Although several reviews have been published in the discussed field, e.g. [63], their scope remains too broad to address in sufficient depth the performance of the methods and results. The reviewers, contrary to us, often do not stick to methodological guidelines [52,64–66].

6. Conclusions

In this study, we provided a broad overview of papers published on applying artificial intelligence to emotion and affect recognition using cardiovascular signals. There is a heterogeneous group of recognized emotions, used artificial intelligence models and explored datasets. Due to poor reporting, the included studies lack essential information in the description of the population, methods and experiment, making it impossible to generalize the results and draw firm conclusions.

If these results are to be used in the future for health monitoring, wellness management and assisted living, more effort should be put into adequately conducting research. The authors should focus more on methodological aspects of their experiments and reporting, including utilizing validated guidelines.

In our future work, we plan to conduct an in-depth analysis of the credibility and the reporting in the field and assess the performance of different ML classifiers used for emotion and affect recognition. We also want to propose a standardized outline for other researchers conducting experiments in emotion and affect elicitation, further aiming at internal states recognition.

Additional Information

Authors' Contributions

We followed CRediT (Contributor Roles Taxonomy) to report on the authors' contributions.

1. Paweł Jemioło (PJ): Conceptualisation, Data Curation, Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation, Visualisation, Writing – Original Draft, Writing – Review & Editing.
2. Dawid Storman (DS): Conceptualisation, Formal analysis, Methodology, Investigation, Validation, Visualisation, Writing – Original Draft, Writing – Review & Editing.
3. Maria Mamica (MM): Conceptualization, Investigation, Writing – Original Draft, Writing – Review & Editing.
4. Mateusz Szymkowski (MS): Conceptualization, Investigation, Writing – Original Draft, Writing – Review & Editing.
5. Patryk Orzechowski (PO): Conceptualization, Investigation, Writing – Original Draft.

Acknowledgement

We thank Weronika Dranka (WD) for supporting with data extraction.

Conflict of Interest

The authors declare that they have no conflict of interest.

Funding Sources

Research project supported by program *Excellence initiative -- research university* for the University of Science and Technology.

References

- [1] R. W. Picard, *Affective Computing*. MIT Press, 1997.
- [2] T. Thanapattheerakul, K. Mao, J. Amoranto, and J. H. Chan, "Emotion in a century: A review of emotion recognition," in *Proceedings of the 10th International Conference on Advances in Information Technology*, pp. 1–8, 2018.
- [3] P. Ekman, W. V. Friesen, M. O'sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti, et al., "Universals and cultural differences in the judgments of facial expressions of emotion," *Journal of personality and social psychology*, vol. 53, no. 4, p. 712, 1987.
- [4] J. A. Russell and L. F. Barrett, "Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant," *Journal of personality and social psychology*, vol. 76, no. 5, p. 805, 1999.
- [5] L. F. Barrett, *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt, 2017.
- [6] A. Wierzbicka, "Talking about emotions: Semantics, culture, and cognition," *Cognition & Emotion*, vol. 6, no. 3-4, pp. 285–319, 1992.
- [7] S. Sosnowski, A. Bittermann, K. Kuhnlenz, and M. Buss, "Design and evaluation of emotion-display eddie," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3113–3118, IEEE, 2006.
- [8] M. Malfaz and M. A. Salichs, "A new architecture for autonomous robots based on emotions," *IFAC Proceedings Volumes*, vol. 37, no. 8, pp. 805–809, 2004.
- [9] T. Ogata and S. Sugano, "Emotional communication between humans and the autonomous robot which has the emotion model," in *Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No. 99CH36288C)*, vol. 4, pp. 3177–3182, IEEE, 1999.
- [10] S. Zepf, J. Hernandez, A. Schmitt, W. Minker, and R. W. Picard, "Driver emotion recognition for intelligent vehicles: A survey," *ACM Computing Surveys (CSUR)*, vol. 53, no. 3, pp. 1–30, 2020.
- [11] C. L. Lisetti and F. Nasoz, "Affective intelligent car interfaces with emotion recognition," in *Proceedings of 11th International Conference on Human Computer Interaction, Las Vegas, NV, USA, 2005*.
- [12] T. Souto, H. Silva, A. Leite, A. Baptista, C. Queirós, and A. Marques, "Facial emotion recognition: Virtual reality program for facial emotion recognition—a trial program targeted at individuals with schizophrenia," *Rehabilitation Counseling Bulletin*, vol. 63, no. 2, pp. 79–90, 2020.
- [13] M. Steppan, L. Fürer, N. Schenk, K. Schmeck, et al., "Machine learning facial emotion recognition in psychotherapy research. a useful approach?," 2020.
- [14] M. P. Ketelaars, A. Mol, H. Swaab, S. van Rijn, et al., "Emotion recognition and alexithymia in high functioning females with autism spectrum disorder," *Research in Autism Spectrum Disorders*, vol. 21, pp. 51–60, 2016.
- [15] R. Matin and D. Valles, "A speech emotion recognition solution-based on support vector machine for children with autism spectrum disorder to help identify human emotions," in *2020 Intermountain Engineering, Technology and Computing (IETC)*, pp. 1–6, 2020.
- [16] L. Pepa, M. Capecci, and M. G. Ceravolo, "Smartwatch based emotion recognition in parkinson's disease," in *2019 IEEE 23rd International Symposium on Consumer Technologies (ISCT)*, pp. 23–24, 2019.
- [17] M.-C. Lee, S.-Y. Chiang, S.-C. Yeh, and T.-F. Wen, "Study on emotion recognition and companion chatbot using deep neural network," *Multimedia Tools and Applications*, vol. 79, no. 27, pp. 19629–19657, 2020.
- [18] M. Karna, D. S. Juliet, and R. C. Joy, "Deep learning based text emotion recognition for chatbot applications," in *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*, pp. 988–993, IEEE, 2020.
- [19] F. Guo, W. L. Liu, Y. Cao, F. T. Liu, and M. L. Li, "Optimization design of a webpage based on kansei engineering," *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 26, no. 1, pp. 110–126, 2016.
- [20] J. Fleureau, P. Guillotel, and Q. Huynh-Thu, "Physiological-based affect event detector for entertainment video applications," *IEEE Transactions on Affective Computing*, vol. 3, no. 3, pp. 379–385, 2012.
- [21] P. Jemioło, B. Giżycka, and G. J. Nalepa, "Prototypes of arcade games enabling affective interaction," in *International Conference on Artificial Intelligence and Soft Computing*, pp. 553–563, Springer, 2019.
- [22] S. Li and W. Deng, "Deep facial expression recognition: A survey," *IEEE Transactions on Affective Computing*, 2020.
- [23] M. B. Akçay and K. Oğuz, "Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers," *Speech Communication*, vol. 116, pp. 56–76, 2020.
- [24] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "Eeg-based emotion recognition using deep learning network with principal component based covariate shift adaptation," *The Scientific World Journal*, vol. 2014, 2014.
- [25] F. Agraftoti, D. Hatzinakos, and A. K. Anderson, "Ecg pattern analysis for emotion detection," *IEEE Transactions on affective computing*, vol. 3, no. 1, pp. 102–115, 2011.
- [26] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu, and X. Yang, "A review of emotion recognition using physiological signals," *Sensors*, vol. 18, no. 7, p. 2074, 2018.
- [27] M. Dumas, "Emotional expression recognition using support vector machines," in *Proceedings of international conference on multimodal interfaces*, Citeseer, 2001.
- [28] Y.-L. Lin and G. Wei, "Speech emotion recognition based on hmm and svm," in *2005 international conference on machine learning and cybernetics*, vol. 8, pp. 4898–4901, IEEE, 2005.
- [29] M. Li, H. Xu, X. Liu, and S. Lu, "Emotion recognition from multichannel eeg signals using k-nearest neighbor classification," *Technology and health care*, vol. 26, no. S1, pp. 509–519, 2018.

- [30] L. Li, Y. Zhao, D. Jiang, Y. Zhang, F. Wang, I. Gonzalez, E. Valentin, and H. Sahli, "Hybrid deep neural network–hidden markov model (dnn-hmm) based speech emotion recognition," in *2013 Humaine association conference on affective computing and intelligent interaction*, pp. 312–317, IEEE, 2013.
- [31] J. Vanschoren, J. N. van Rijn, B. Bischl, and L. Torgo, "Openml: Networked science in machine learning," *SIGKDD Explorations*, vol. 15, no. 2, pp. 49–60, 2013.
- [32] R. S. Olson, W. La Cava, P. Orzechowski, R. J. Urbanowicz, and J. H. Moore, "Pmlb: a large benchmark suite for machine learning evaluation and comparison," *BioData Mining*, vol. 10, p. 36, Dec 2017.
- [33] B. K. Beaulieu-Jones and C. S. Greene, "Reproducibility of computational workflows is automated using continuous analysis," *Nature biotechnology*, vol. 35, no. 4, pp. 342–346, 2017.
- [34] M. Mamica, P. Kapłon, and P. Jemioło, "Eeg-based emotion recognition using convolutional neural networks," in *International Conference on Conceptual Structures, ICCS*, Springer, 2021.
- [35] F. P. George, I. Mannafee, *et al.*, "Recognition of emotional states using eeg signals based on time-frequency analysis and svm classifier," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, p. 1012, 04 2019.
- [36] M. Nuijten, C. Hartgerink, M. Assen, *et al.*, "The prevalence of statistical reporting errors in psychology (1985-2013)," *Behavior research methods*, vol. 48, 10 2015.
- [37] M. Bakker and J. Wicherts, "The (mis)reporting of statistical results in psychology," *Behavior research methods*, vol. 43, pp. 666–78, 04 2011.
- [38] Y. Ling, T. An, L. W. Yap, B. Zhu, S. Gong, and W. Cheng, "Disruptive, soft, wearable sensors," *Advanced Materials*, vol. 32, no. 18, p. 1904664, 2020.
- [39] P. Jemioło, D. Storman, M. Mamica, M. Szymkowski, P. Orzechowski, and W. Dranka, "Emotion recognition from cardiovascular signals using artificial intelligence – a systematic review," March 2021. Available from: <https://osf.io/kzj8y/>. Accessed on September 19, 2021.
- [40] P. Jemioło, D. Storman, M. Mamica, M. Szymkowski, P. Orzechowski, and W. Dranka, "Emotion recognition from cardiovascular signals using artificial intelligence – a systematic review (registration)," March 2021. Available from: <https://osf.io/nj7ut/>. Accessed on September 19, 2021.
- [41] K. H. Kim, S. W. Bang, and S. R. Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Medical and biological engineering and computing*, vol. 42, no. 3, pp. 419–427, 2004.
- [42] A. Konar, A. Halder, and A. Chakraborty, "Introduction to emotion recognition," *Emotion Recognition: A Pattern Analysis Approach*, pp. 1–45, 2015.
- [43] J. McNames and M. Aboy, "Statistical modeling of cardiovascular signals and parameter estimation based on the extended kalman filter," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 1, pp. 119–129, 2007.
- [44] J. F. Knight and C. Baber, "A tool to assess the comfort of wearable computers," *Human factors*, vol. 47, no. 1, pp. 77–91, 2005.
- [45] B. Copeland, "Artificial intelligence: Definition, examples, and applications," *Encyclopedia Britannica*, 2020.
- [46] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (eeg) classification tasks: a review," *Journal of neural engineering*, vol. 16, no. 3, p. 031001, 2019.
- [47] A. Botchkarev, "Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology," *arXiv preprint arXiv:1809.03006*, 2018.
- [48] M. Ouzzani, H. Hammady, Z. Fedorowicz, and A. Elmagarmid, "Rayyan—a web and mobile app for systematic reviews," *Systematic reviews*, vol. 5, no. 1, p. 210, 2016.
- [49] P. F. Whiting, A. W. Rutjes, M. E. Westwood, S. Mallett, J. J. Deeks, J. B. Reitsma, M. M. Leeflang, J. A. Sterne, and P. M. Bossuyt, "Quadas-2: a revised tool for the quality assessment of diagnostic accuracy studies," *Annals of internal medicine*, vol. 155, no. 8, pp. 529–536, 2011.
- [50] R. F. Wolff, K. G. Moons, R. D. Riley, P. F. Whiting, M. Westwood, G. S. Collins, J. B. Reitsma, J. Kleijnen, and S. Mallett, "Probast: a tool to assess the risk of bias and applicability of prediction model studies," *Annals of internal medicine*, vol. 170, no. 1, pp. 51–58, 2019.
- [51] M. J. Benton, A. M. Hutchins, and J. J. Dawes, "Effect of menstrual cycle on resting metabolism: A systematic review and meta-analysis," *PloS one*, vol. 15, no. 7, p. e0236025, 2020.
- [52] M. D. McInnes, D. Moher, B. D. Thombs, T. A. McGrath, P. M. Bossuyt, T. Clifford, J. F. Cohen, J. J. Deeks, C. Gatsonis, L. Hooft, *et al.*, "Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the prisma-dta statement," *Jama*, vol. 319, no. 4, pp. 388–396, 2018.
- [53] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 18–31, 2011.
- [54] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 42–55, 2011.
- [55] J. A. M. Correa, M. K. Abadi, N. Sebe, and I. Patras, "Amigos: A dataset for affect, personality and mood research on individuals and groups," *IEEE Transactions on Affective Computing*, 2018.
- [56] 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 journal of biomedical and health informatics*, vol. 22, no. 1, pp. 98–107, 2017.
- [57] M. Egger, M. Ley, and S. Hanke, "Emotion recognition from physiological signal analysis: a review," *Electronic Notes in Theoretical Computer Science*, vol. 343, pp. 35–55, 2019.
- [58] R. E. Haamer, E. Rusadze, I. Lsi, T. Ahmed, S. Escalera, and G. Anbarjafari, "Review on emotion recognition databases," *Hum. Robot Interact. Theor. Appl.*, vol. 3, pp. 39–63, 2017.

- [59] S. Saganowski, P. Kazienko, M. Dziezyc, P. Jakimow, J. Komoszynska, W. Michalska, A. Dutkowiak, A. Polak, A. Dziadek, and M. Ujma, "Consumer wearables and affective computing for wellbeing support," *MobiQuitous 2020 - 17th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, Dec 2020.
- [60] X. Yang, L. Tan, and L. He, "A robust least squares support vector machine for regression and classification with noise," *Neurocomputing*, vol. 140, pp. 41–52, 2014.
- [61] R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar, and T. Alhussain, "Speech emotion recognition using deep learning techniques: A review," *IEEE Access*, vol. 7, pp. 117327–117345, 2019.
- [62] W.-L. Zheng, J.-Y. Zhu, Y. Peng, and B.-L. Lu, "Eeg-based emotion classification using deep belief networks," in *2014 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, IEEE, 2014.
- [63] A. Saxena, A. Khanna, and D. Gupta, "Emotion recognition and detection methods: A comprehensive survey," *Journal of Artificial Intelligence and Systems*, vol. 2, no. 1, pp. 53–79, 2020.
- [64] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Prisma 2009 flow diagram," *The PRISMA statement*, vol. 6, p. 97, 2009.
- [65] A. Liberati, D. Altman, J. Tetzlaff, C. Mulrow, P. Gøtzsche, J. Ioannidis, M. Clarke, and P. Devereaux, "The prisma statement for reporting systematic and meta-analyses of studies that evaluate interventions," *PLoS Medicine*, vol. 6, no. 7, pp. 1–28, 2009.
- [66] J. P. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, and V. A. Welch, *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons, 2019.