Robotic Emotion Monitoring for Mental Health Applications: Preliminary Outcomes of a Survey

Marat Rostov[1], Md Zakir Hossain[1, 2 [0000-0003-1892-831X]], Jessica Sharmin Rahman[1]

[1] Research School of Computing, Australian National University (ANU), Canberra, Australia
[2] Research School of Biology, ANU & Agriculture and Food, The Commonwealth Scientific and Industrial Research Organization (CSIRO), Black Mountain, Canberra, Australia

Abstract. Maintaining mental health is crucial for emotional, psychological, and social well-being. Currently, however, societal mental health is at an all-time low. Robots have already proven useful in medicine, and robot assisted mental therapies through emotional monitoring have great potential. This paper reviews 60 recent papers to determine how accurately robots can classify human emotions using the latest sensor technologies. Among 18 different signals, it was determined that EDA sensors are best for this application. Our findings also show that CNN outperforms SVM, SVR, KNN and LDA for classifying EDA data with an average of 79% accuracy. This is further improved with the addition of RGB sensor data.

Keywords: emotion recognition, machine learning, physiology, robots, sensors

1 Introduction

In 2017, Twenge et al. [1] surveyed 600,000 people living in the U.S for exploring societal mental health. Their goal was to determine self-reported depression rates amongst the current youth. During 2009-2017, they found the proportion of adolescents reporting a depressive episode had doubled. These results were a sobering reminder of the times we live in. Social isolation has been linked to a plethora of mental issues [2], [3]. This is especially a problem, when developed countries are finding that the number of single-person households is increasing. In some countries these make up 60% of residences [4]. Loneliness is not limited to millennials either, an aging and busier population is resulting in our elders having less familial support. Hence, our society is one of great loneliness and mental health is sure to degrade if nothing is done.

We believe that robots will play a crucial role in maintaining our mental health. Not by replacing practicing professionals, but as an aid to improve mental therapies. Assistive robots have already seen widespread success in healthcare [5], [6]. In addition, facial expression analysis found useful for detecting doubt effect, subjective beliefs, differentiating between observed real and posed smiles, and assessing mental health [7], [8]. Thus, it is possible to use robots for monitoring mental health of patients and report them back to a professional therapist. Currently, detailed mental health monitoring is done through patients recounting their days. However, people often face difficulties remembering events accurately. Robots can help us achieve this through emotion monitoring. As an example, sadness is found highly correlated with depression for clinical diagnoses [9]. As such, recognizing emotions through robots could provide an objective account of patients' mental states. It is worthwhile to note that without AI or machine learning (ML) methods, robots must use a hard-coded program to achieve emotion recognition. Thus, our aim is to determine which non-invasive physiological signals and ML methods are most appropriate for emotion monitoring through robots.

2 Method

We searched six academic databases, namely IEEE Xplore, Google Scholar, University SuperSearch, Scopus, Pubmed Central, and ResearchGate. Three keywords were derived in consultation with University librarians: **robot***, **emotion recognition**, and **sensor*** (where * denotes wildcard characters). The key words cover the area related to robotic, emotional, and physiological contexts where 'sensor' term is used to find related physiological based papers. The search results were kept between 1 June 2015 to 31 May 2020 for finding most up-to-date papers. If there were more than 200 results (Google Scholar, University SuperSearch, and Scopus), the results were sorted by the engine's definition of relevance for accumulating most relevant papers of interest, and the top 200 results were added to the screening pool. This resulted in a collection of 805 papers in total, of which 619 were unique.

Then, articles were excluded if they were either: 1) publications that were not original peer-reviewed papers, 2) papers not related to emotion recognition, 3) papers that do not mention applicability of research to improving robots or machines, or 4) papers that do not state research applicability in a mental therapy context. For validation of excluded results, two reviewers acted independently on the 619 unique papers. There were 17% mismatched papers from each reviewer. By discussing with third reviewer, 60 papers were finally included for detail analyses.

3 Results

We found 18 different signals were studied in 60 included papers. They are categorized into *physiological* and *physical* signals[1]. Fig. 1 (left) shows the distribution of sensors used in experiments. In total, 112 sensors were studied across the 60 papers. The heart signal category makes up the largest fraction at 28.6%. RGB was the most common signal studied across the 60 papers (15.2%). The RGB is followed by EDA

[1] *Physiological*: BVP, PPG, HR, ECG, EDA (electrodermal activity), EEG, EOG, ST (skin conductance), HDR (hemodynamic response), motion (accelerometers and positional information), resp. (respiratory signals)

Physical: 3D, RGB, Text, env. (environmental), radio signal, SA (speech audio), Tactile (pressure sensor)

(14.3%), EEG (13.4%), and others. Overall, the physiological and the physical signals make up 70.5% and 29.5% of the total signals studied respectively.

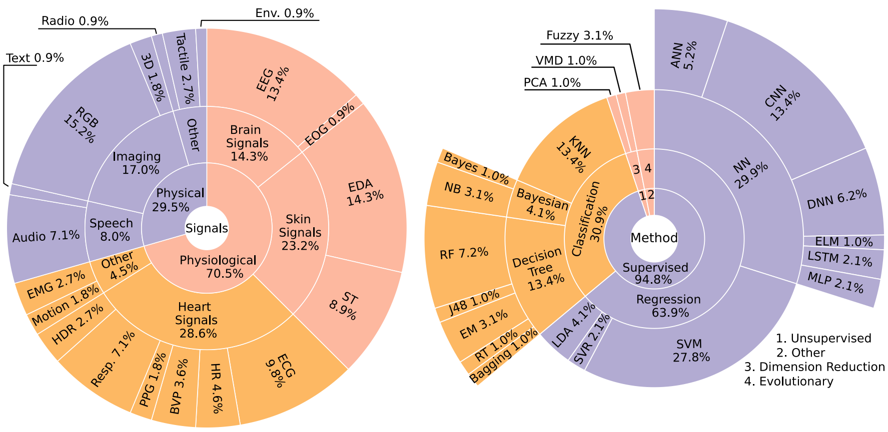

Fig. 1. Signals[1] studied (left) and used ML[2] methods (right) in the included papers.

The distribution of used ML[2] classifiers is displayed in Fig. 1 (right). The classifiers have been categorized into supervised, unsupervised, and others. The supervised methods are subcategorized into classification and regression. It has been seen that SVM is a highly representative method with 27.8% usability followed by KNN (13.4%) and CNN (13.4%). Methods involving artificial neural networks accounted for 29.9% cases. The overwhelming majority, 94.8%, of classifiers are supervised. Meanwhile dimensionality reduction methods were the only unsupervised methods used, at 2.1% usability.

4 Discussion and Conclusion

Emotion classification from facial images or video data appeared six times in the top 15 papers with highest classification accuracies, for example [10], [11]. However, top three highest scoring scores reported from four or less sample sizes. EEG appeared twice in the top 15 papers with an average accuracy of 92.0% between the two. HR, EDA, ECG, body temperature, motion data, speech audio, EMG, BVP were also appeared in the list of top 15 papers. The 'top 15 papers' represents the highest accuracies among the included 60 papers. This is highly important as it means that the majority of commonly studied non-invasive physiological signals have the potential to be very accurate in emotion recognition.

[2] *ML methods*: ANN (artificial neural network), Bagging, Bayes, LDA, KNN, PCA, VMD (variational mode decomposition), MLP, CNN (convolutional neural network), DNN (deep neural network), ELM (extreme learning machine), LSTM, NB, SVM, SVR, Fuzzy, RF (random forest), J48 (J48 decision tree), EM (ensemble methods), RT (regression tree).

The highest scoring method was found CNN, applied on RGB data. However, Fig. 1 (right) shows that the most common classifier was SVM (27.8%). We believe there are two reasons for the low number of papers for CNN. Firstly, training CNNs or any deep networks require huge data. This is hard to come from physiological signals, and even the largest sample size in the included 60 papers was 457 for an EDA experiment. However, image sets can have potentially thousands of faces, and this does not count video datasets. Computational effort could be another factor as SVM is faster than CNNs.

Initially, RGB cameras appeared to be the best single sensor for emotion recognition. EEG, ECG, and other physiological sensors are also found viable for high accuracy emotion recognition. However, these fall short on comfort and ease of use due to longer setup times. Future work should analyses the sensors long term comfort and ease of implementation. This way, we will have a better idea of which sensors are better implementable in the mental therapy application.

EDA was determined to be best suited towards mental health monitoring robots. Out of the 60 papers, 14 papers considered some form of EDA in emotion recognition. Of these, only two used standalone EDA [12], [13], while others used a fusion of sensors. It has been found that even standalone EDA has great potential for emotion recognition tasks. Thus, an inbuilt EDA sensor [14], [15] on robot can be used for the tasks. Furthermore, CNN-classified EDA provided an average of 79.0% accuracy over six papers. Hence, CNN appears to be the best classifier for EDA data if the sample size is sufficient to allow it. In addition, robot would have a microphone and RGB camera to pick up speech and facial expressions. The fusion of EDA, RGB and speech audio would provide accuracy well above standalone CNN classified EDA data of 79.0%. Overall, we found which sensors we can use in robots to assist in mental health monitoring. Evaluating the accessibility, accuracy, and applicability on 18 unique sensors over 60 papers, EDA was the single-best sensor for this application. Our findings also show that CNN performs better than other ML classifiers.

This research helps pave the way towards improved social robotics and robot assisted mental health solutions. We have shown that using new classifiers with even standalone EDA provides accurate and objective emotion recognition. But in combination with camera and microphone data, we can attain very high emotion recognition accuracy, as required for clinical depression diagnosis. Minimizing the number of sensors required allows cost effective robots with less failure points. Hence, a robot with a camera and microphone feed providing support to a patient wearing an EDA wristband could be the emotion monitoring robot that helps to keep us all in good mental health.

Due to the lack of space in here, our future work will explore the overall process to conduct the literature review as proposed by Kitchenham [16], which includes, but not limited to, detailed planning, execution and data synthesis, the sample size used, the gender of the participants, suitable EDA sensors, RGB data collection and classification, different research dimensions, and broad future challenges.

References

1. J. M. Twenge, T. E. Joiner, M. L. Rogers, and G. N. Martin, "Increases in Depressive Symptoms, Suicide-Related Outcomes, and Suicide Rates Among U.S. Adolescents After 2010 and Links to Increased New Media Screen Time", *Clinical Psychological Science,* vol. 6, no. 1, pp. 3-17, 2017.
2. R. Mushtaq, S. Shoib, T. Shah, and S. Mushtaq, "Relationship between loneliness, psychiatric disorders and physical health? A review on the psychological aspects of loneliness", *J Clin Diagn Res,* vol. 8, no. 9, pp. WE01-4, Sep 2014.
3. M. E. Loades *et al.*, "Rapid Systematic Review: The Impact of Social Isolation and Loneliness on the Mental Health of Children and Adolescents in the Context of COVID-19", *J Am Acad Child Adolesc Psychiatry,* 2020.
4. K. D. M. Snell, "The rise of living alone and loneliness in history", *Social History,* vol. 42, no. 1, pp. 2-28, 2017.
5. M. Simon. "This Incredible Hospital Robot Is Saving Lives. Also, I Hate It". WIRED. 2015. Available: https://bit.ly/34JHZxR [Accessed by October 2020].
6. C. W. R. University. "5 Medical Robots Making a Difference in Healthcare" Case Western Reserve University. 2020. Available: https://bit.ly/3iggbZP [Accessed by October 2020].
7. X. Zhu, Z. Qin, T. Gedeon, R. Jones, M. Z. Hossain, S. Caldwell, "Detecting the Doubt Effect and Subjective Beliefs Using Neural Networks and Observers' Pupillary Responses", *Int. Conf. on Neural Information Processing,* LNCS, vol. 11304, pp. 610-621, 2018.
8. M. Z. Hossain, T. Gedeon, and R. Sankaranarayana, "Using temporal features of observers' physiological measures to distinguish between genuine and fake smiles". *IEEE Transactions on Affective Computing*, vol. 11, no. 1, pp.163-173, 2018.
9. S. Mouchet-Mages and F. J. Baylé, "Sadness as an integral part of depression", (in eng), *Dialogues Clin Neurosci,* vol. 10, no. 3, pp. 321-7, 2008.
10. Ruiz-Garcia, M. Elshaw, A. Altahhan, and V. Palade. "Deep learning for emotion recognition in faces", *Lecture Notes in Computer Science*, vol. 9887, pp. 38-46, 2016.
11. J. Yang, R. Wang, X. Guan, M. M. Hassan, A. Almogren, and A. Alsanad, "AI-enabled emotion-aware robot: The fusion of smart clothing, edge clouds and robotics", *Future Generation Computer Systems,* vol. 102, pp. 701-709, 2020.
12. L. Boccanfuso *et al.*, "A thermal emotion classifier for improved human-robot interaction", *25th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 718-723, 2016.
13. G. Yin *et al.*, "User Independent Emotion Recognition with Residual Signal-Image Network", *International Conference on Image Processing,* pp. 3277-3281, 2019.
14. FitBit. "Fitbit Debuts Sense, Its Most Advanced Health Smartwatch; World's First with EDA Sensor for Stress Management, Plus ECG App, SpO2 and Skin Temperature Sensors". FitBit. 2020.
15. M. Z. Hossain, T. Gedeon, R. Sankaranarayana, "Observer's Galvanic Skin Response for Discriminating Real from Fake Smiles", *27th Australasian Conference on Information Systems*, pp. 1-8, 2016.
16. B. Kitchenham, "Procedures for performing systematic reviews". *Keele University,* vol. 33, pp. 1–26, 2004.