Assessing Basic Emotion via Machine Learning: Comparative Analysis of Number of Basic Emotions and Algorithms

Caryn Vowles
Mechanical and Materials Engineering
Queen's University
Kingston, Canada
16cjv@queensu.ca

Mackenzie Collins
Mechanical and Materials Engineering
Queen's University
Kingston, Canada
15mc103@queensu.ca

Dr. T. Claire Davies
Mechanical and Materials Engineering
Queen's University
Kingston, Canada
claire.davies@queensu.ca

2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) | 979-8-3503-7149-9/24/$31.00 ©2024 Authors | DOI: 10.1109/EMBC53108.2024.10782053

Abstract— **This paper explores the use of machine learning (ML) methods to identify "clusters" of basic emotions based on pleasure, arousal, and dominance (PAD). The data was obtained from the Dataset for Emotion Analysis using Physiological Signals (DEAP), data collected within the Building and Designing Assistive Technology (BDAT) Lab using the International Affective Picture System (IAPS), and the scores of PAD from the IAPS. The objective is to develop an algorithm that maps a PAD score to clusters that express emotions, e.g., sadness or happiness. The elbow method was used to determine the optimal number of clusters (4-8), and nine different ML algorithms were compared. Decision Trees, polynomial support vector machines (SVMs) and linear SVMs provided accurate results. The Decision Tree demonstrated efficiency, during both testing and validation, in identifying the same clusters when analyzing both the DEAP and IAPS datasets. The dataset included limited data for each emotion creating the possibility of overfitting. However, when evaluating the results relative to previous research, the results added to the understanding of the nuances of emotion self-reporting and modelling.**

Keywords—emotions, PAD, affective computing, machine learning

I. INTRODUCTION

Recent advances in affective computing have led to an increased interest in the study of human emotions. Emotions are important in shaping human behaviour, communication, and social interactions. Understanding the underlying mechanisms that govern the experience and expression of emotions is crucial for developing intelligent algorithms that can recognize, respond, and simulate emotions in a natural and empathetic way [1]. Though every day we may use a variety of words to explain our emotions: ecstatic, ashamed, judgemental, enraged, inadequate and astonished, there are more frequently used words to describe the base state of emotions: awe, amusement, happy, excitement, content, fear, sadness, disgust, anger [2]. Ekman considered our basic emotions to be Fear, Anger, Happiness, Sadness, Disgust, and Surprise [1]. These basic emotions are considered discrete and do not account for mixed emotions.

Researchers have developed various self-reporting methods, such as the Positive and Negative Affect Schedule (PANAS) and Pleasure-Arousal-Dominance (PAD) theory, to help understand an individual's meaning of emotional words. PANAS is a self-report questionnaire used to measure an individual's current level of positive and negative affect [3]. However, the PAD theory attempts to better understand the nuances of emotion when self-reported. It is more robust than the PANAS in that it describes human emotions using three fundamental dimensions: pleasure, arousal, and dominance [4]. These dimensions are critical in understanding the complexity of human emotions. To measure emotions on the PAD scale, researchers use the Self-Assessment Manikin (SAM). The SAM consists of three different Likert scales that pictorially represent the dimensions. Each dimension's scale ratings range from 1 to 9. This visual tool allows participants to rate their feelings using non-verbal descriptive images [5]. Though PAD scores are reported on a continuous scale, Likert responses are chosen by self-report. Prior to understanding the nuances of using PAD scores to interpret mixed emotions, the mapping to independent basic emotions must be better understood.

Eliminating the influence of the third dimension, dominance, is widespread when using PAD scores for affective computing and artificial intelligence to describe emotions from stimuli. While many researchers remove dominance from the computational representation of affect, recent literature has indicated that dominance is essential when modelling emotions [6]. Constantinescu *et al.* [7] also proposed that dominance should be included when using the PAD method for clustering emotions.

Minimal evidence can be found in the literature that maps PAD scores to commonly used emotion identifying terms (e.g. happy or sad). The objective of the current research is to use the PAD scores to identify clusters of basic emotions within the PAD three-dimensional space using machine learning (ML) methods. These results aim to support the development of an algorithm that can efficiently map a PAD score to an emotion-identifier.

II. METHODS

The DEAP (Dataset for Emotion Analysis using Physiological Signals) provides the PAD scores obtained in response to video images [8]. To the best of our knowledge, there is a gap in current literature that connects these PAD scores to emotion-identifying terms used in everyday expression of emotion. Using the DEAP data, a computational model was created that mapped PAD scores to clusters that represent basic emotions. The current research aims to evaluate various machine learning models to identify clusters of similar PAD scores which may then be linked to an emotion identifying term.

To validate this model, data was collected from participants who viewed pictures from the International Affective Picture

System (IAPS) and reported their emotions via the SAM. The IAPS is a standardized system that contains a variety of images across a wide range of semantic categories. PAD scores have been collected and analyzed for each individual image within the IAPS [9]. Using DEAP, and the collected data using IAPS, various computational models were developed to connect PAD scores to clusters of identified emotions.

A. DEAP Dataset

In the current study, only the online subjective annotation component of the DEAP dataset was used (Table I) [8]. This dataset was initially created by having participants view 40 one-minute music video clips online. The participants were asked to record their emotional responses according to three categories: 1) PAD on the SAM; 2) a self-identification of emotion from a list of 16 options, including Pride, Elation, Joy, Satisfaction, Relief, Hope, Interest, Surprise, Sadness, Fear, Shame, Guilt, Envy, Disgust, Contempt, and Anger; 3) the strength of the selected emotion on a scale of 1-4, where level 1 is weak, and level 4 is strong. Of the 120 samples, each one-minute music video clip was viewed by a minimum of 14 and a maximum of 16 participants, with the participants being different in each trial. Koelstra *et al.*[8] provided a comprehensive discussion in their paper on the initial stimuli selection of the music video clips.

TABLE I. DEAP SUMMARY ADAPTED FROM[8]

Online Subjective Annotation	
Stimuli	1 minute affective videos (60 selected via affective tags and 60 manually)
Number of Ratings	14-16 /video
Outputs	Pleasure, Arousal, Dominance, Emotion, Emotion Strength
Pleasure, Arousal and Dominance Scores	Discrete scale 1-9
Emotions	16 Emotion Description Words
Strength of Emotion	Scale 1-4

B. Data Collection

Participants (BDAT Dataset): Thirty-one participants (9-25 years) with the mean age of 21.8 years and a standard deviation of 3.8 years. who were typically developing (no disabilities) participated in this research. The study was approved by the Queen's University Health Sciences and Affiliated Teaching Hospitals Research Ethics Board (HSREB) and each individual provided informed consent or assent (depending on age).

Stimuli: The International Affective Picture System (IAPS) was used to elicit emotions. Since this research is part of a larger project seeking to identify the emotions of children, the IAPS deck for children aged 7-9 years was used. The researchers removed sensitive images from the study, leaving 52 images that were approved by the HSREB.

Protocol: The protocol involved participants viewing 30 pictures. Before each image, a grey slide with the text "Get ready to rate the next image..." was displayed for five seconds to normalize emotions to baseline. Then, one image from the deck was shown for six seconds. Participants were asked to identify how they felt on each of the pleasure, arousal, and dominance scales (i.e. the three dimensions of the SAM) by pointing to the appropriate image on the screen. A researcher observed and recorded the corresponding PAD scores based on the SAM image reported. Additional details regarding the study are given by Collins *et al.* [10].

C. Computational model

The objective of the computational model (Fig. 1) was to connect the PAD scale to clusters that represented emotions. Though PAD scores have been heavily used in research, they do not directly connect to specific emotional terms used in everyday communication and are an unlabeled dataset [7]. To create these emotion labels of PAD scores, the number of potential basic emotions had to be identified; this was done via K-means clustering and the elbow method on the DEAP dataset. The elbow method is a powerful technique used to determine the optimal number of clusters when applying K-means clustering. This method involves plotting the variance explained by the clusters against the number of clusters and identifying the point at which the curve begins to level off or form an elbow shape. The value at which the curve reaches an asymptote signifies the ideal number of clusters to retain for the analysis. By leveraging the elbow method with the DEAP dataset, it was possible to find the optimal number of clusters that best represented emotions. Then using K-means clustering the PAD scores were labelled by cluster number, each representing a basic emotion.

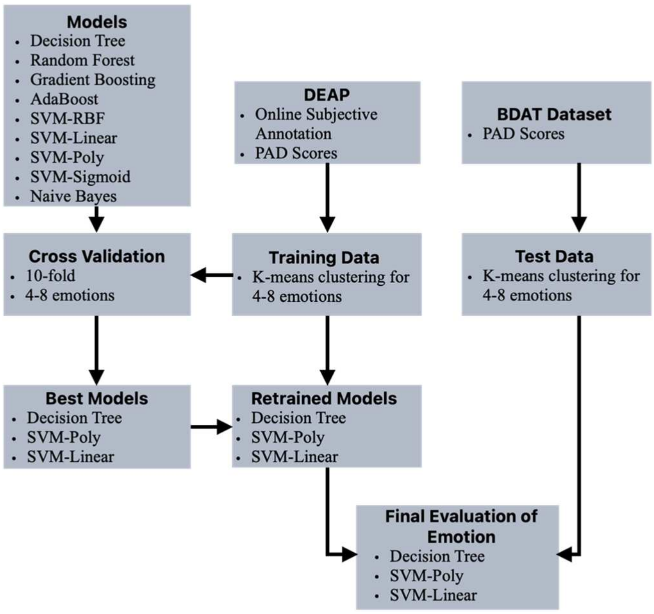

Fig. 1. Flowchart of Methodology

The next step was to identify a machine learning model that could sort new PAD scores (datapoints) into these emotion clusters. Nine machine learning models were investigated, each using 4-8 clusters. The machine learning models included various trees (Decision, Random Forest, Gradient Boosting Machines, AdaBoost), support vector machines (RBF, linear, poly, sigmoid) and a naïve Bayes model. Decision trees are a machine learning model that sort datapoints using a defined sequence of rules. In a Decision Tree, each node represents a prediction that has been made based on features (measurable pieces) of the data, and the branches represent the possible outcomes. A Random Forest model combines multiple decision trees. Each tree is trained on a random subset of the data and features, and the final prediction is made by taking the average predictions of all the trees. In a Random Forest model, each of the trees are given equal weight in determining the final prediction. While Gradient Boosting Machines and Adaboost are similar to Random Forest, they differ in that each tree within the forest makes a weighted contribution to the final prediction, meaning some trees will have a greater impact on the final prediction than others. Support vector machines (SVMs) are a type of model that searches for the optimal hyperplane to separate the data into different classes. The algorithm maps the data to a higher-dimensional space and identifies the hyperplane that maximizes the margins separating each class. Different kernel functions (RBF, linear, poly, sigmoid) are used to map the data to different spaces. Naïve Bayes is a probabilistic model that uses Bayes' theorem to predict the probability of a class given conditionally independent features. These models were trained and tested with a ten-fold cross-validation method on the DEAP dataset.

Finally, speed and accuracy of each of the machine learning models were compared. The three top-performing models; Decision Trees, poly SVM, and linear SVM, were validated with the BDAT Lab data and the IAPS data scores on each cluster. Since the stimuli between the two datasets were the same, the predicted emotion cluster identified by the participant's data could be directly compared to the predicted emotion cluster from the original IAPS PAD scores.

III. RESULTS

This study used machine learning techniques to identify the optimal number of basic emotion categories that can be differentiated within the PAD model of emotion. Various algorithms were tested and compared on the basis of accuracy to determine the optimal model. The elbow method was used to map the number of clusters relative to the sum of the squared distance/inertia up to 20 clusters and the elbow occurred at approximately 5 clusters. However, to minimize loss of potential data, a range of potential clusters (4-8) was evaluated. An example of data represented as six clusters in the PAD space is shown in Fig.2.

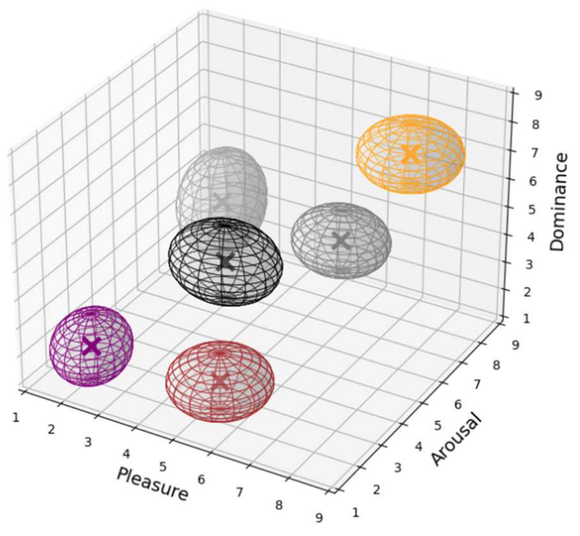

Fig. 2. Six Clusters of Emotion

Using ten-fold cross-validation, nine different machine learning algorithms were tested for each of 4-8 clusters of basic emotions using the DEAP dataset. Results for the Decision Tree **six-cluster model** at each of the ten-folds are shown in Table II to provide an example.

TABLE II. DECISION TREE TEN-FOLD RESULTS FOR SIX CLUSTERS

Fold	Score(%)	Score time ($x10^{-4}$s)
0	99.3902	3.33
1	100	1.92
2	99.3902	1.56
3	98.1595	1.61
4	99.3865	1.67
5	98.773	1.51
6	98.1595	1.48
7	100	1.45
8	100	1.43
9	100	1.45

*Scores are given as percent accuracies.

Table II indicates that the model is 100% accurate 4 out of 10 times during training with a minimum accuracy of 98.1%, and the slowest scoring time was 3.33x 10^{-4}s. This ten-fold cross-validation analysis of the DEAP dataset was done for each number of clusters between four and eight.

An example of the results for the **six-cluster model** for each of the algorithms are found in Table III. Results are color-coded to indicate performance. Green represents both the highest accuracy scores and the fastest times, orange and yellow indicate decrease in performance respectively, while those highlighted in red represent both the lowest accuracy scores and the slowest times.

TABLE III. ANALYSIS OF MODELS FOR 6 CLUSTERS

Algorithm		Timing		
	Score (%)	*Fit time ($x10^{-4}$)*	*Score time ($x10^{-4}$)*	*Total time ($x10^{-4}$)*
Decision Tree	99.33	10.07	1.74	11.81
Random Forest	99.45	1293.45	125.32	1418.76
Gradient Boosting	99.51	5112.97	16.19	5129.16
AdaBoost	63.14	459.24	38.83	498.06
SVM-RBF	98.28	213.40	64.83	278.23
SVM-linear	98.59	59.85	10.09	69.94
SVM-poly	99.02	91.20	7.16	98.36
SVM-sigmoid	21.44	2027.86	164.31	2192.18
Bayes	96.39	7.06	2.74	9.80

The highest performing machine learning model for each number of clusters was chosen by examining both score and time. For example, for the six-cluster model given in Table III, the most accurate algorithm is the Gradient Boosting. However, it is also one of the slowest. Bayes is the fastest algorithm for both fitting and scoring, but its accuracy is the third from the poorest. The selected algorithm for the six-clusters models is a decision tree, which had the fastest scoring time and was within 0.5% accuracy of the most accurate algorithm, followed by the polynomial and linear

SVMs due to their accuracy scores being both greater than 98% with fast score times.

The stimuli for all data were the same (IAPS pictures). The predicted emotions resulting from the 890 PAD scores collected in the BDAT lab were compared to the predicted emotions of the PAD scores from the IAPS for each stimulus. Results in Table V indicate that the six-cluster-decision tree model had the most consistent results when comparing both datasets resulting in 47% similarity. All other models with different cluster numbers resulted in a lack of conformance among the two datasets with similarity scores of less than 13%. The best outcome, resulting in similarity scores of 47%, included the six cluster versions of the Decision Tree and the linear SVM. Based on this analysis, it appears that six clusters best represent base emotions from these two datasets. Decision Tree and linear SVM algorithms were the most accurate in the testing phase, however, the Decision Tree algorithm takes less time in the validation phase.

TABLE IV. SIMILARITY OF MODELS PREDICTIONS FOR IAPS VS OUR COLLECTED DATA

Cluster	**Decision Tree**		**Poly SVM**		**Linear SVM**	
	Same	***Different***	***Same***	***Differe nt***	***Same***	***Different***
4	***111*** *(12%)*	***779*** *(88%)*	***113*** *(13%)*	***777*** *(87%)*	***122*** *(12%)*	***768*** *(86%)*
5	***86*** *(10%)*	***804*** *(90%)*	***87*** *(10%)*	***803*** *(90%)*	***92*** *(10%)*	***798*** *(90%)*
6	***416*** *(47%)*	***474*** *(53%)*	***384*** *(43%)*	***506*** *(57%)*	***416*** *(47%)*	***474*** *(53%)*
7	***88*** *(10%)*	***802*** *(90%)*	***95*** *(11%)*	***795*** *(89%)*	***86*** *(10%)*	***804*** *(90%)*
8	***67*** *(8%)*	***823*** *(92%)*	*74 (8%)*	***816*** *(92%)*	*71 (8%)*	***819*** *(92%)*

IV. DISCUSSION

Our results indicate that the best model to use in the assessment of emotions is the 6-cluster Decision Tree model. However, since the average accuracy of the testing phase was greater than 98% and the validation phase only reached 47% correct, it is possible that overfitting may have occurred.

The IAPS data does not link words that express emotion to the images that are shown for each trial, but only PAD scores for that image. Since the current study followed the IAPS protocol explicitly, words to express emotion were also not requested from the data collection within the BDAT Lab. Although a comparison was drawn between the two datasets to numerically identify the number of clusters that were present in the data as determined by the PAD scores, there was no link between the numerical results and any words that the individual may have used to express their emotions. The DEAP dataset, however, did link these PAD scores to emotion words. A previous paper explored the relationships between PAD scores and words used to express emotion [6]. Although six clusters were also identified in that paper, only four could be labelled as explicit emotions (Happy, Anger, Disgust, and Sadness). Evidence from these two papers and others [7, 11] suggests that there are six clusters of emotions that result when using the data from PAD scores. Ekman [1] argues that there are six basic emotions (Fear, Anger, Happiness, Sadness, Disgust, and Surprise), which suggests that the unlabelled clusters may be Fear and Surprise, but insufficient evidence exists to draw this conclusion. Images within the IAPS to evoke Fear and Surprise were limited within the data collection at the BDAT Lab as a result of receiving ethics approval only for images that could also be shown to children between 7-9 years (as this research is part of a larger study).

Mickels *et al.* [2] asked participants in an open-ended style which emotion was evoked when observing the negative and the positive IAPS images. Using a calculation that expressed the percentage of frequency of a specific emotion label relative to the total number of negative emotion labels, Mickels *et al.* identified eight basic emotion words (awe, amusement, happy, excitement, content, fear, sadness, disgust, anger). Mickels found that pictures often invoked more than one emotion and would be considered as "mixed emotion" representations. In the current study, only discrete emotions that were identified through independent clusters were considered, minimizing the complexities of mixed emotions. Future work should consider removing images with reported mixed emotion from the Mickels list to see if the identification of clusters results in a higher similarity score. The lack of ability to detect independent clusters that include mixed emotions is a limitation of the current work and, although complex, should be considered in future research. The researchers believe that identification of independent emotions should occur prior to attempting the classification of mixed emotions.

A systematic review by Branco *et al.* [12] that evaluated the use of the IAPS within a laboratory setting to elicit real-life emotions found differences in PAD scores due to age (young vs old), sex (Male vs Female) and culture. This suggests a need to develop participant specific models when attempting to identify an emotion of an individual. Since the current work used the DEAP dataset, it was not possible to build participant-specific models that would take those differences into account. A greater number of participants would improve the accuracy when selecting the basic emotions. Validating the model's performance with larger and more diverse datasets would strengthen the reliability and applicability of the findings across different populations and contexts.

V. CONCLUSION

In conclusion, the present study investigated the optimal number of clusters for analyzing emotion PAD scores and identified the most accurate algorithm for clustering basic emotions. The elbow method was used to determine the optimal number of clusters to sort basic emotions and it was determined that the number of clusters ranges from four to eight. The testing phase suggested that the Decision Tree, poly SVM, and linear SVM algorithms were among the top three algorithms. During validation of these models using IAPS original study and data collected in the BDAT lab, six basic clusters of emotions appear to exist that are best identified using an SVM poly model. Future work could consider mixed emotions over the discrete basic emotions discussed in this paper.

Limitations of this study include the presence of mixed emotions in some images that may decrease accuracy during the testing phase, and the inability to build participant specific models since the DEAP dataset did not have sufficient data per participant to make this feasible.

Overall, the findings of this study have important implications for researchers seeking to relate PAD scores to commonly used vocabulary for the expression of emotions.

ACKNOWLEDGMENT

The authors wish to thank members of the BDAT lab, past and present, who helped with this work. This work was supported in part by OGS (Ontario Graduate Scholarship), NSERC [RGPIN-2016- 04669] and NSERC CREATE READi.

REFERENCES

[1] P. Ekman, "Universal Facial Expressions of Emotion," *California Mental Health Research Digest,* vol. 8, no. 4, 1970.

[2] J. A. Mikels, G. R. Fredrickson Bl Fau - Larkin, C. M. Larkin Gr Fau - Lindberg, S. J. Lindberg Cm Fau - Maglio, P. A. Maglio Sj Fau - Reuter-Lorenz, and P. A. Reuter-Lorenz, "Emotional category data on images from the International Affective Picture System," (in eng), no. 1554-351X (Print).

[3] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: The PANAS scales," *Journal of Personality and Social Psychology,* vol. 54, pp. 1063-1070, 1988, doi: 10.1037/0022-3514.54.6.1063.

[4] I. Bakker, T. van der Voordt, P. Vink, and J. de Boon, "Pleasure, Arousal, Dominance: Mehrabian and Russell revisited," *Current Psychology,* vol. 33, no. 3, pp. 405-421, 2014/09/01 2014, doi: 10.1007/s12144-014-9219-4.

[5] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry,* vol. 25, no. 1, pp. 49-59, 1994/03/01/ 1994, doi: https://doi.org/10.1016/0005-7916(94)90063-9.

[6] C. Vowles and C. Davies, "Emotion Based on the PAD Theory of Emotion: A Comparison of DEAP and ANEW Datasets," *Pre-print* 2023.

[7] A. C. Constantinescu, M. Wolters, A. Moore, and S. E. MacPherson, "A cluster-based approach to selecting representative stimuli from the International Affective Picture System (IAPS) database," *Behavior Research Methods,* vol. 49, no. 3, pp. 896-912, 2017/06/01 2017, doi: 10.3758/s13428-016-0750-0.

[8] S. Koelstra *et al.*, "DEAP: A Database for Emotion Analysis ;Using Physiological Signals," *IEEE Transactions on Affective Computing,* vol. 3, no. 1, pp. 18-31, 2012, doi: 10.1109/t-affc.2011.15.

[9] M. M. Bradley and P. J. Lang, "The International Affective Picture System (IAPS) in the study of emotion and attention," in *Handbook of emotion elicitation and assessment.*, (Series in affective science. New York, NY, US: Oxford University Press, 2007, pp. 29-46.

[10] M. Collins and T. C. Davies, "Emotion differentiation through features of eye-tracking and pupil diameter for monitoring well-being," (in eng), no. 2694-0604 (Electronic).

[11] P. EKman. "Universal Emotions." Paul Ekman Group. https://www.paulekman.com/universal-emotions/ (accessed June 2, 2023).

[12] D. A.-O. Branco, F. A.-O. Gonçalves Ó, and S. A.-O. Badia, "A Systematic Review of International Affective Picture System (IAPS) around the World. LID - 10.3390/s23083866 [doi] LID - 3866," (in eng), no. 1424-8220 (Electronic).