# Introducing Affective Computing

Dane Wanke

COMP 3190

Dr. John Anderson

University of Manitoba

Winnipeg, Manitoba

April 5, 2023

Outline

# Introducing Affective Computing

## 1. Abstract

Affective computing is a new branch of artificial intelligence. Still in its infant stages, there remains plenty to research. This paper explores what it is affecting computing is and aims to achieve. Next, the concept of classification and regression machine learning algorithms at the core of affective computing are introduced which is then followed by a walkthrough of a visual-based affective computing system. After covering each stage and providing concrete examples, alternative systems that rely upon other methods of input besides visuals, data selection and testing methods are discussed. Lastly, examples of problems that need to be addressed are given, of which, numerous original solutions are proposed.

## 2. What is Affective Computing?

Affective computing is the subfield of artificial intelligence that aims to provide computers with the processing capabilities to evaluate and display emotion. In their processing, computers lack contextual awareness required to correctly address many issues. Moreover, “today's evidence indicates that a healthy balance of emotions is integral to intelligence, and to creative and flexible problem solving” [Picard, 2000]. As one such example, consider an online platform used for education. While such a platform provides value, it may be less effective at engaging students, identifying subtleties from questions asked or noticing when a student is confused or may benefit from a momentary pause. Therefore, a system which could detect confusion, adapt content to improve engagement and generate answers based on contextual details from questions could prove useful. More generally, this is one of many issues that affective computing aims to address by offering new ways to interact and benefit from technology in a more natural approach to how people behave.

## 3. Classification and Regression

Classification and regression are two methods machine learning, a core component of affective computing, uses to derive its predictions. Classification deals with qualitative data, that matches on similarities as opposed to regression which depends upon numerical values. In other words, classification is used with discrete data whereas regression handles continuous data [Harrington, 2012]. Although both can be used in affective computing models, each has its use case. Regression is faster and easier to work with [Harrington, 2012] while classification algorithms like support vectors machines or $k^{th}$ nearest neighbor, also known as SVM and k-NN respectively, in the context of affective computing, are seen as more effective and thus tend to be used more frequently.

## 4. Algorithmic Flow

*Input/Pre-Processing Stage*

Most affective computing systems process visuals, which certainly introduces many challenges. Therefore, due to its popularity and complexity it is the implementation focused in the following example of an affective computing system. To

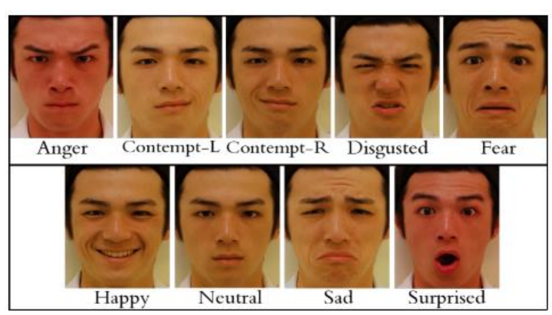



Figure 1: An example of expressions that have been cropped [Najah, 2017]

start, this system requires either images or video as input which, once received, searches for a face [Najah, 2017]. Once located within the image or video frame, it is then cropped and resized which consequently reduces image sizes and increases processing speeds. The resulting image is then converted to greyscale and has its contrast increased to exaggerate differences and facial characteristics [Najah, 2017].

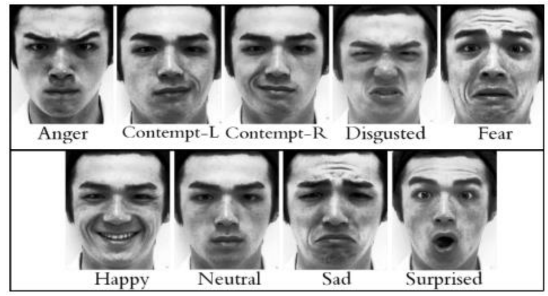



Figure 2: An example of the same faces as figure 1 after greyscale conversion and contrast is increased [Najah, 2017]

*Feature Extraction Stage*

Next, we enter the feature extraction stage. The purpose of this stage is to alter an image into arrangements of micropatterns for further processing [Turabzadeh et al., 2018]. The arrangement of micropatterns are clear divisions in the input images size where the smaller these divisions are the better, since each individual portion is used to identify a single facial feature [Ahonen et al., 2006]. This further illustrates the importance of image cropping and resizing. Since each resulting portion becomes the same pre-defined size, it is important for each image to be of the same size and common facial features to fall into similar segments. In otherwards, an effective visual-based system has the same facial features reappear in the same subsections consistently. Local Binary Pattern Histogram, also known as LBP, is a texture defining algorithm which is then applied as the next step. For each portion of the

Figure 3: an example of image division [Ahonen et al., 2006]

image, LBP processes each pixel further dividing the image into temporary blocks of nine per pixel [Ahonen et al., 2006]. That is, given a pixel, we consider its surrounding neighbors. The original value of the center pixel is treated as a threshold value which is then compared to the value of each adjacent pixel. These comparisons are recorded by binary digits 1 and 0 which will depend upon if the adjacent pixel had an intensity greater than or equal to the threshold [Ahonen et al., 2006].

Figure 4: an input image (left) turned LBP (right) [Turabzadeh et al., 2018].

After each comparison is made, LBP uses these findings to sum powers of two so that the new encoded value, which will replace the previous that was used as the threshold [Ahonen et al., 2006]. It is worth noting this new value is based on the relationship between the center pixel and its neighboring points [Najah, 2017]. Starting at the top right corner pixel, the binary digit we add, assuming that its value satisfies the threshold is $2^0$, the one below $2^1$ which continues in a clockwise fashion.

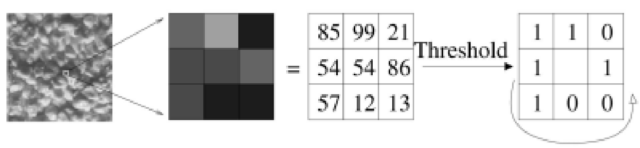
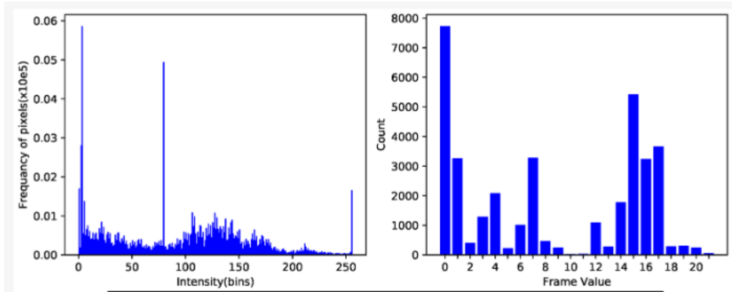



Figure 5: the center pixels new intensity value is being calculated based on its neighbors [Ahonen et al., 2006]

This trend repeats itself until the last pixel, the one immediately above the center is selected which potentially adds $2^7$ [Ahonen et al., 2006]. Again, the point of this is to sum all these powers of two to identify a new intensity value for the center pixel that ranges between 0 and 255. After every computation is handled, it is time to convert the resulting values into their respective histograms [Turabzadeh et al., 2018]. This is done for each divided portion of the image, these histograms are then concatenated together thus acting as a descriptor for the entire image. Interestingly enough each of the individual histograms represent a "calculated feature from the photo as per its division and that in the spatially enhanced histogram, we effectively have a description of the face on three different levels of locality: The LBP labels for the histogram contain information about the patterns on a pixel-level, the labels are summed over a small region to produce information on a regional level, and the regional histograms are concatenated to build a global description of the face" [Turabzadeh et al., 2018].




| Value | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Count | 7727 | 3266 | 404 | 1294 | 2085 | 227 | 1018 | 3278 | 463 | 247 | 37 |
| Value | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
| Count | 45 | 1097 | 284 | 1778 | 5429 | 3241 | 3666 | 294 | 310 | 245 | 58 |

Figure 6: the histogram descriptor for an arbitrary processed image [Turabzadeh et al., 2018].

*Classification Stage*

Finally, it is time for the classification stage. As previously mentioned, affective computing systems typically employ classification algorithms such as k-NN and SVM, both of which will be illustrated. K-NN compares Euclidian distances from the processed input, the summed histogram vector, against each points in the model in search of the k closest points for some pre-defined arbitrary value k > 0 [Najah, 2017]. The input is then matched to the same category as that which the majority of these k points shared. The resulting class is then outputted as the prediction from the system [Najah,

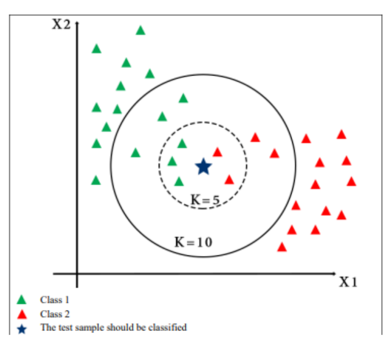



Figure 7: an example of k-NN in action [Najah, 2017]

2017]. Usually, k is a small positive integer. For example, with k=1, the test data is

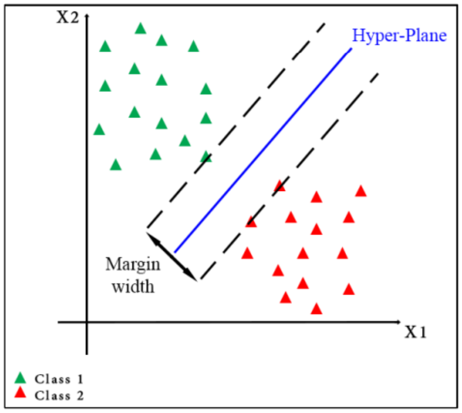



Figure 8: an example of SVM in action [Najah, 2017]

allocated to the class of that single nearest neighbor [Najah, 2017]. However, the overall consistency of k seems to vary as per where some studies had their model improved as k increased [Turabzadeh et al., 2018]. Whereas k-NN searches for a number of closest points, SVM establishes boundaries, also known as separating hyper-planes, around each class for points to fall within. Furthermore, this algorithm also uses a margin to detect outliers within the model [Zhang et al., 2016]. This is extremely useful in the sense that it helps with generalization and prevents overfitting so that input is less affected from any biases and extreme cases. Since traditionally “SVM was developed for two-class problem and more than two emotions exist, techniques such as winner-takes-all (WTA) can break the task into multiple 2-class tasks” [Zhang et al., 2016] for proper processing. Under these conditions, SVM generally output more accurate predictions while k-NN requires less computation time.

## 5. Other Forms of Input

Besides visual based systems, audio based affective computing systems are also quite popular. Mostly, they tend to focus on what was said to infer how the speaker felt, however, more recently researchers have shifted their focus into how these messages are delivered [Picard, 2000]. That is, different emotions have auditory cues that can be derived from regular speech. This alternative of using audio rather then

| | Fear | Anger | Sadness | Happiness | Disgust |
|---|---|---|---|---|---|
| Speech rate | much faster | slightly faster | slightly slower | faster or slower | very much slower |
| Pitch average | very much higher | very much higher | slightly lower | much higher | very much lower |
| Pitch range | much wider | much wider | slightly narrower | much wider | slightly wider |
| Intensity | normal | higher | lower | higher | lower |
| Voice quality | irregular voicing | breathy chest tone | resonant | breathy blaring | grumbled chest tone |
| Pitch changes | normal | abrupt on stressed syllables | downward inflections | smooth upward inflections | wide down-ward terminal inflections |
| Articulation | precise | tense | slurring | normal | normal |

Figure 9: examples of auditory cues for their respective emotion [Picard, 2000]

visuals presents its own challenges as well in the sense that language structure can be difficult to classify and that each individual has their own distinct vocal ranges that need to be known in advance. In addition to these higher-level signals, Rosalind Picard also states the following low-level signals may be used: muscle response and nervous system activity, blood volume pressure, galvanic skin response, pupillary dilation, skin conductance and color, temperature and respiratory patterns. All of which are becoming easier to track with unintrusive, wearable devices [Picard, 2000]. Our body strongly corelates feelings to physical states, so much so, that to some degree emotional state can be influenced by forcing characteristics such as a smile or sweaty palms [Picard, 2000]. Moreover, "in his 1862 thesis, Duchenne identified independent expressive muscles in the face, such as the muscle of attention, muscle of lust, muscle of disdain or doubt, and muscle of joy" [Picard, 2000] which was later used to identify direct mappings between muscles and emotional space [Picard, 2000]. Therefore, although less popular, these lower-level systems are very capable, even by themselves, maybe even more so as since they can also capture more discrete cues.

## 6. Testing Effectiveness

Like most machine learning programs, in affective computing "there are two main source of faults in machine learning programs: the data and the model" [Braiek and Khomh, 2020]. Also like most machine learning programs, "the amount of data to be able to converge and make meaningful inferences, can make data collection a challenging step" [Braiek and Khomh, 2020] especially considering the cultural and physical differences across populations per individual and as they express emotion. Therefore, it follows that these distinct points are needed for the creation of more robust systems. Furthermore, dirty data has been consistently shown to have a negative impact on performance [Braiek and Khomh, 2020] so data should also be cleaned as its collected and processed to ensure data consistency such as image sizing and respective formats. Later, generated model effectiveness can be measured with cross validation methods such as k-fold, Monte Carlo and leave-one-out comparisons. With which, unintuitively, random data sets are preferred over those deliberately chosen [Ramezan et al., 2019]. With k-fold the data set in question is divide into an arbitrary number of unique portions and use each resulting subset, except one, to test run tests and see of the excluded data, which generate output as expected. This process generally repeats itself until every portion of data is tested. Next, all averages are summed thus returning the models overall effectiveness [Ramezan et al., 2019]. Leave-one-out comparisons are similar except that in addition, one data set excluded entirely. The advantage of this method is that irrelevant data may show itself as once removed, the model's accuracy remains unaffected [Ramezan et al., 2019]. Finally, Monte Carlo runs multiple iterations of tests in which data is divided randomly per run and the results are averaged [Ramezan et al., 2019].

## 7. Complexities of Emotion

### *Keeping Record*

The difficulty of classifying emotional patterns is often what drives research in affective computing [Turabzadeh et al., 2018]. An emotional state has multiple components including both cognitive and physiological awareness and each individuals subjective feelings [Picard, 2000]. Furthermore, across all forms, emotions are time-invariant, inconsistent and influenced by both physical and internal states [Picard, 2000]. That is, there are no guarantees for a particular emotion or reaction for any circumstance even considering only a single individual. This may be due to the effects of emotional saturation where experiencing the same emotion can either result in diminishing returns or intensify with each recurring trigger [Picard, 2000]. Alternatively, it is also possible for a situations to cause a reaction for some, while none for others [Picard, 2000]. This suggests that future systems could potentially be improved by maintaining records and history of previous emotional states.

### *Processing Multiple Signals*

Another issue that could be better addressed is one suggested by Rosalind Picard herself. This is the inability to detect individual authenticity [Picard, 2000]. That is, to know that how a person truly feels is how they present themselves. This is an area higher level systems that use visual or audio-based input may struggle against, however, could be addressed with the addition of lower-level sensors. As a primary component in lie detection systems, heartbeats have proven themselves to detect inconsistencies. However, a lesser-known fact is that heartbeats share very intimate details about an individual. For example, when one person loves another, their heartbeat changes to synchronize with that of the other persons. A heartbeat will also speed up when excited and slow down when calm, thus simplifying emotion recognition considerably regarding Picard's two-dimensional model of emotions centered around "arousal (calm/excited), and valence (negative/positive)" [Picard, 2000]. Perhaps more generally, while there is currently no system that has 100% accuracy, processing multiple signals represented by multiple forms of input seems promising. This does however have its own challenges such as when using visual and audio processing together since speaking requires the mouth to move, which also varies per person, thus making it more difficult to capture facial expressions while speaking" [Picard, 2000].

*Multi Label Classification*

Currently, the range of detectable emotions is limited to anger, contempt, disgust, fear, happiness, neutrality, sadness and excitement [Picard, 2000]. However, besides these pure emotions there are also more complex emotions that are mixed. That is, these emotions are formed by a combination of the 8 above, or multiple may alternate so frequently that it gives the impression of being only a single emotion. This is why multi label classification should also be explored. That is, to identify these combinations and complex situations in which a displayed emotion may appear fuzzy in the sense that it could belong to more than one category. However, with the use of multiple diverse inputs and multiple k-NN these results can be further analyzed and thus derive potential edge cases, and provide less obvious situations with multiple observations for further analysis, for example, against lower-level signals.

*Addressing Speed*

One of the strengths both SVM and k-NN share is that they are relatively fast algorithms. This is especially important considering affective computing's biggest challenge is speed. Currently, there is no systems that can run in real time, however computing resources such as graphics cards and specialized central processing units may change this. A graphic cards processing power, called GU-GPA, speeds up the affective computing "processes 80-fold thus allowing new readings every 15 seconds" [Turabzadeh et al., 2018] whereas FPGA is a specialized central processing unit that goes "20x faster than GU-GPU" [Turabzadeh et al., 2018]. FPGA also has overall better performance especially in regards to artificial intelligence, they are easy to customize and modify, low-cost, reliability and require little long-term maintenance [Turabzadeh et al., 2018] which seems promising for the future of affective computing.

## 8. Conclusion

In conclusion, affective computing will drastically change the landscape of computing, smoothing out interactions and providing solutions that better address human nature. This is a shared purpose amongst all variations of such a system which may appear as any combination of audio, visual and lower-level signal-based systems. Inside of which, data is pre-processed, extracted then classified against a data driven model, which can be improved with the help of the various tests and data conscientious decisions. Lastly, while there has yet to be shown that a perfect system exists, it is possible that the solution could be found in that more data is needed on the unique relationships between different emotional states and their characteristics. Focusing on knowing more about how the body reacts overall, not just in relation to a single input, in order to effectively discriminate emotions accurately over a more complex spectrum. In addition, further research should be conducted into the roles lower-level signals, saved historical emotional states and the use of different models could provide.

# 9. Reference List

[Ahonen et al., 2006] Ahonen, T., Hadid, A. and Pietikainen, M. “Face Description with Local Binary Patterns: Application to Face Recognition”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12), pp.2037–2041, doi:https://doi.org/10.1109/tpami.2006.244, 2006.

[Braiek and Khomh, 2020] Braiek, H.B. and Khomh, F. “On testing machine learning programs”. *Journal of Systems and Software*, [online] 164, p.110542. doi:https://doi.org/10.1016/j.jss.2020.110542, 2020.

[Harrington, 2012] Harrington, Peter, *Machine learning in action*. Manning Publications., 2012.

[Najah, 2017] Goma Najah. “*EMOTION ESTIMATION FROM FACIAL IMAGES”*. In IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, [online] , 2017.

[Picard, 2000] Picard, Rosalind, *Affective computing*. The MIT Press., 2000.

[Ramezan et al., 2019] A. Ramezan, C., A. Warner, T. and E. Maxwell, A. “Evaluation of Sampling and Cross-Validation Tuning Strategies for Regional-Scale Machine Learning Classification”. *Remote Sensing*, [online] 11(2), p.185. doi:https://doi.org/10.3390/rs11020185, 2019.

[Turabzadeh et al., 2018] Turabzadeh, S., Meng, H., Swash, R., Pleva, M. and Juhar, J. “Facial Expression Emotion Detection for Real-Time Embedded Systems”. In *Technologies*, [online] 6(1), p.17., doi:https://doi.org/10.3390/technologies6010017, 2018.

[Zhang et al., 2016] Zhang, Y.-D., Yang, Z.-J., Lu, H.-M., Zhou, X.-X., Phillips, P., Liu, Q.-M. and Wang, S.-H. “Facial Emotion Recognition Based on Biorthogonal Wavelet Entropy, Fuzzy Support Vector Machine, and Stratified Cross Validation”. *IEEE Access*, [online] 4, pp.8375–8385. doi:https://doi.org/10.1109/access.2016.2628407, 2016.