An embodied cognition classifier of human emotion and senses

D. Kernot [a]

[a] Defence Science and Technology, Edinburgh, SA, Australia.
Email: david.kernot@dst.defence.gov.au

Abstract: Drawing on the idea of embodied cognition, where the human body and its environment influence the way a person thinks and feels, we test if it is possible to classify suicide attackers from their writing by using words affected by the emotional and modality-specific systems in the brain. We use their final notes and manifestos compared with normal bloggers' posts to train a machine learning classifier. We compare two support vector machine classifier models using linear and radial base function kernels. In this exploratory study, receiver operating characteristic curves show encouraging separation accuracies. These models offer an 11-13% improvement over methods using only emotion or sense categories. This study supports the idea that an embodied cognition classifier better discriminates the way a person thinks and feels rather than treating the body and mind as separate entities and may help in reflecting behavior and applying influence in online social systems.

Keywords: *Machine learning, vector support machines, suicide terrorism, embodied cognition*

1.   INTRODUCTION

Classifying people into one group or another has many uses today. Being able to classify people into emotional and cognitive categories is beneficial for a number of reasons, such as mental health (Varga, 2018; Gjelsvik, Lovric, & Williams, 2018), profiling people for high-pressure employment categories such as police recruitment, being a surgeon, or a highly effective sportsperson (Nieuwenhuys & Oudejans, 2017), and for controlling hate speech in online forums (Alorainy, 2018). However, it has its down side. Recently, bots have been used to influence individuals in online social systems and sway voting outcomes using negative emotional content (Stella, Ferrara, & De Domenico, 2018). As highlighted in the recent US Presidential elections, persuasive communication is particularly effective when tailored to people's unique psychological characteristics and motivations (Matz et al., 2018). Social influence can be successfully persuasive when synchronization between communicators and receivers is aligned to the experiences, thoughts and emotions of others, as highlighted by embodied social cognition (Falk & Scholz, 2018). Emotions are intertwined with human thinking and behaviour, not psychologically distinct from them, and can alter our cognitions and thereby transform our personal and social world, exerting a powerful influence over a range of judgments and ethical decisions (Drodge & Murphy, 2002).

Embodied cognition is grounded in cognitive neuroscience and psychology, and research into it has risen exponentially over the last 25 years (Gjelsvik, Lovric, & Williams, 2018). Embodied cognition regards the human body and the environment as significant factors in the way we think and feel (Guell, Gabrieli, & Schmahmann, 2018). This is done by processing both emotional and modality-specific systems in the brain (Barsalou et al, 2003; Niedenthal et al, 2005; Mahon, 2015; Tillman, & Louwerse, 2018), where both emotion and the sensory multi-modal specific processing of memory work together (Niedenthal, 2007; Dreyer & Pulvermüller, 2018), also known as semantic cognition (Ralph et al, 2017:2).

Embedded cognition is grounded in the idea that the body is critical in idea generation and then acting on those ideas, or thoughts. When objects and events are viewed through the eyes of the self they typically become emotionally coloured, and thereby more intimately related to one's sense of self (Northoff et al., 2006: 441). The senses are critical in memory recall and creating new ideas. We draw on Cechetto and Weishaupt (2016: pp84-92) to describe the biological process of thinking and idea creation within the brain. This is where inputs from the body's visual, auditory, somatosensory, somatomotor, gustatory and olfactory sensations are processed in their individual cortices and channelled through the entorhinal cortex (EC) region, which acts as a central hub, or a multimodal association area for sensory information processing. In the creation of new episodic memories, to encode our daily personal experiences, and to retrieve episodic memories, information flows along these sensory-specific and temporal isocortical brain regions to the hippocampus. It is well understood that the hippocampus is associated with memory and in particular long-term memory, and is a part of the limbic system that regulates emotion before sensations reach the dentate gyrus (DG), a region in the hippocampus responsible for the formation of new episodic memories.

Physiological changes have long been associated with emotion (Healey, 2014). Emotion modulates almost all modes of human communication – word choice, tone of voice, facial expression, gestural behaviours, posture, skin temperature and clamminess, respiration, muscle tension, and more (Picard, Vyzas, & Healey, 2001). But the processing of emotions from words and the processing of emotions from pictures or faces share the same neurophysiological mechanisms (Herbert et al, 2018). Emotional perception may be driven by sensory information stored in memory (Doyle & Lindquist, 2018). There is compelling evidence that emotional content conveyed by abstract symbols such as words can elicit consciously retrievable affective feeling states (Herbert et al, 2018).

To test the idea that embodied cognition can better characterize an individual using psychological characteristics, we draw on the processing of emotional and modality-specific systems in the brain to model it using a machine learning classifier, and test if an approach that describes embodied cognition is a better classifier over separate instances of either emotion or sensory processing.

1.1   Machine learning classifiers

There are many types of machine learning classifiers that have been used to extract emotion and human behaviours from text. Machine learning classifiers have been used to decode brain stimuli, mental states, and behaviours in individuals (Pereira, Mitchell, & Botvinick, 2009). Deep neural networks of sensory information have been constructed using speech recordings and video to predict emotion from subjects (Tzirakis, 2017). Decision trees and Naïve Bayes classifiers can classify human sentiment, extricating positive or negative polarities from social media text (Singh, Singh, & Singh, 2017). Convolutional neural network and Support Vector Machine (SVM) classifiers can map abstract concepts such as semantic category, writing style, or sentiment from text documents (Arras et al., 2018). The use of SVMs is quite common, and they have been

used to extract human sentiment and emotions from social media data (Bisio, Oneto, & Cambria, 2017; Rani, & Singh, 2017; Rout et al, 2018; Seyeditabari, Tabari, & Zadrozny, 2018).

1.2   Support vector machines

SVMs are a form of pattern recognition that have their basis in R.A. Fisher's (1936) linear discriminator function and the construction of decision surfaces where an optimal hyperplane maps the maximal margin between the vectors of two classes using a small amount of training data known as support vectors to determine this margin (Cortes & Vapnik, 1995). The SVM process has two stages where the dataset is trained to obtain an optimal model, and then the model is used as a classifier to predict information (Chang & Lin, 2011).

Using an existing anonymised dataset of two distinct groups, a group of suicide attackers and normal bloggers, we construct a machine learning classifier that can distinguish suicide attackers from normal bloggers using words rich in emotion and sensory information, and used to identify unknown groups of individuals.

2.   METHODS

We used an existing anonymised dataset (Kernot, Bossomaier, & Bradbury, 2017b) that comprises 35 normal bloggers and 25 suicide notes and final manifestos from attacks in the USA, Germany, Canada, Brazil, and Finland. Data were processed with the Linguistic Inquiry and Word Count (LIWC) tool (Pennebaker *et al.*, 2015) and we use the negative emotion and anger sentiment tags that are known to differentiate the writing of normal students from suicide attackers (Egnoto & Griffin, 2016). It was also processed for human senses using RPAS, a technique that identifies self from neurolinguistics indicators (Kernot, Bossamier, & Bradbury, 2018). RPAS is a multidisciplinary approach that accounts for cognitive processing of the senses in the sensory cortex, and for the influence of a person's mental state. In our earlier study we classified the data using LIWC for anger and negative emotion sentiment tags. We used a step-wise multiple regression analysis approach using the Richness, Gender, Referential Activity Power and highly Visual Sensory words. However, Kernot, Bossomaier, & Bradbury (2017a) highlighted the limited role of Gender and we elected to remove it from this study because the word list is also contained within the Referential Activity Power variable. Here, we use Richness, Referential Activity Power and Sensory words that are highly visual. We describe in more detail below how we create a support vector machine to classify our data using ten-fold cross validation from linear and non-linear models and assess its effectiveness using receiver operating characteristic curves.

2.1   Support vector machines

Using the R language and environment for statistical computing (R Core Team, 2017) and an SVM classification package (Meyer et al, 2017), we load the dataset and train and then tune our support vector machine. In the SVM package, there are two methods to train and tune the data. One is a linear kernel and the other is a Gaussian radial basis function kernel. While a radial basis function kernel has been shown to outperform a linear kernel, it has to be tuned correctly using the penalty parameter and the kernel width σ (Keerthi & Lin, 2003) and these are described in the R SVM package as cost and gamma. However, linear classifiers can easily scale up and recent research has shown that for some data sets (e.g., document data such as ours), a linear approach performs as well as kernel classifiers (Huang & Lin, 2016). Therefore, we use 10-fold cross validation and produce the most cost-effective solution after tuning both linear and non-linear (Gaussian radial basis function) kernels from a range of cost and gamma options. We then use the tuned SVM results from each of the linear and non-linear models to calculate prediction probabilities.

2.2   Receiver operating characteristic (ROC) curves

Originally, ROC curves were used to describe radar target detection performance, but now they apply to psychology, medical decision making, bioinformatics, data mining and machine learning (Robin *et al.*, 2011). They are used to depict the tradeoff between hit rates and false alarm rates of classifiers (Metz, 1978). The axes of ROC graphs show sensitivity and specificity. Fawcett (2006) highlights that, in effect, sensitivity and specificity represent two kinds of accuracy: the first for actually positive cases and the second for actually negative cases. He states that one must note carefully that the terms "positive" and "negative" in these definitions concern some particular different state, which must be specified clearly in calculating and quoting sensitivity and specificity values. Metz (1978) refers to sensitivity as the proportion of correctly classified positive observations and specificity as the proportion of correctly classified negative observations, as follows: sensitivity is the true positive rate (TP Rate), also called hit rate, recall or probability of detection. Specificity is the true negative rate (TN Rate), the proportion of negatives that are correctly identified as negatives (see Fawcett, 2006 for a detailed explanation). Generally, the area under the curve (AUC) is a useful metric for the performance of a classifier and is frequently applied for method comparison where a higher AUC means a better

classification (Metz, 1978). In this report, we use the R language and environment for statistical computing, and the Statistical Package for the Social Sciences (SPSS) to generate metrics including ROC curves of the Sensitivity and Specificity values.

3. RESULTS

We create a Support Vector Machine model by combining negative emotion, anger, Richness, Referential Activity Power and Sensory words that are highly Visual (Kernot, Bossomaier, & Bradbury, 2018). We train and then tune our classifiers using ten-fold cross validation against the linear and and radial basis function kernels.

The linear kernel model was scaled and, using 10 fold cross validation, was tested against a general kernel penalizing parameter known as cost, and we used a range of them (0.001. 0.01. 0.1, 1, 5, 10, and 100), resulting in an optimal cost of 1.0. This cost value was provided to the tuned model. The radial basis function kernel model was also scaled, and using 10 fold cross validation, was tested against a range of optional costs (0.001. 0.01. 0.1, 1, 5, 10, and 100). A radial basis function-specific kernel parameter, known as gamma was also selected from a range (0.5, 1, 2, 3, and 4) resulting in an optimal cost of 1.0 and a gamma of 0.5. These values were provided to the tuned model. The predicted classification probabilities that resulted were used and plotted as receiver operating characteristic (ROC) curves to determine the effectiveness of both models to classify the data.

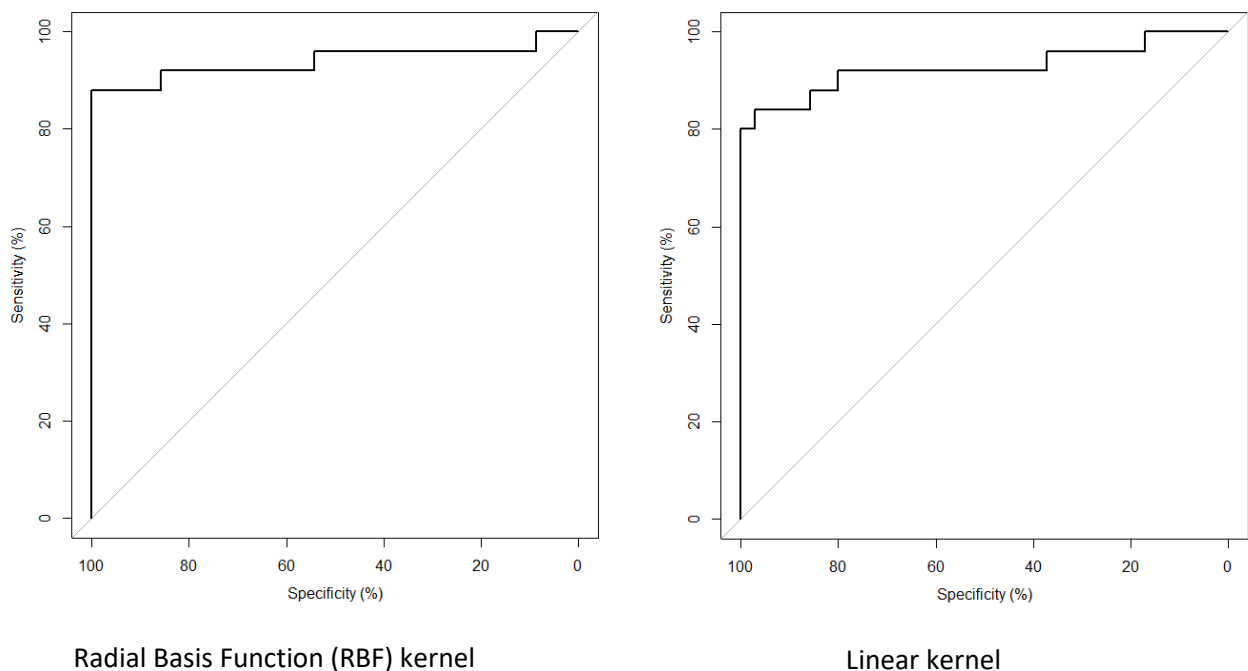Receiver Operating Characteristics (ROC) Curve

Radial Basis Function (RBF) kernel

Linear kernel

Figure 1. ROC curves for the RBF and Linear kernels. Here the RBF kernel tracks the left hand border and then the top border at a better rate than the linear kernel, and we can see a difference between Sensitivity and Specificity. The RBF kernel has less false positives (Specificity) and more true positives (Sensitivity) than the linear kernel.

Using the pROC package in R (Robin et al, 2011), we assessed their effectiveness using ROC curves and calculated the AUC for both kernels. The linear kernel results are encouraging, with an AUC of 0.9269, bounded with an asymptotic 95% Confidence Interval range of 84.47%-100%. When the data is smoothed, by fitting a linear model to the quartiles of the sensitivities and specificities, the binormal AUC is 0.9215. The RBF kernel results are also encouraging. With an AUC of 0.9394, bounded with an asymptotic 95% Confidence Interval range of 85.97%-100%. When the data is smoothed, the binormal AUC is 0.9354. Using the R package pROC, we compare both ROC Curves using DeLong's test for two correlated ROC curves we reject the

hypothesis that the two curves are the same (Z = 0.97846, p-value = 0.3278) and accept there are true differences in AUC's. This is reinforced by a power test (Obuchowski, Lieber, & Wians, 2004) using a significance level = 0.05, the power = 0.4099, offering an alternative two sided solution.

4. DISCUSSION AND CONCLUSION

The concept of classifying individuals using neurolinguistic properties of self from the impacts that sensory and other word types have is new. In our earlier study (Kernot, Bossomaier, & Bradbury, 2017b) using Richness, Gender, referential Activity Power and highly Visual Sensory words, the best result achieved was an Area Under the Curve (AUC) of 94%. We were able to separate the writing of suicide attackers and normal bloggers and achieved an AUC of between 73.4% for negative emotion and 81.1% for anger, compared to an AUC of 83.1% using Richness, Gender, Referential Activity Power and highly Visual Sensory words. While this result was better than using Anger or negative emotion alone, here we have determined that combining both aspects of emotion and the senses achieves a higher result thorough their combination. Here, using the Support Vector Machine model, we use a multidisciplinary approach to characterise people using emotion and thinking. We combine negative emotion, anger, Richness, Referential Activity Power and Sensory words that are highly Visual (see Methods). After we trained and tuned our data, ROC curves highlight improvements over our earlier methods and support the idea of embedded cognition.

Table 1. A comparison of the different methods, showing that the data on human emotion and cognitive and sensory experiences result in a better classification of a person's mindset (in terms of area under the ROC curve) than drawing purely on individual sentiment or cognitive and sensory processes.

Variables	AUC	Type of Human Processing	
Negative Emotion (NE)	0.737	Emotional Sentiment	
Anger	0.811	Emotional Sentiment	
RPAV	0.831	Cognitive and sensory processing	
NE + Anger + RAS (Linear kernel)	0.926	Emotional, cognitive and sensory processing	
NE + Anger + RAS (RBF kernel)	0.939	Emotional, cognitive and sensory processing	

The linear kernel result outperforms the earlier study using separate RPAS by (≈9.5%), anger (by ≈11.5%) and negative emotion (by ≈19.2%) methods. The RBF kernel result outperforms the earlier study using separate RPAS (by ≈10.8%), anger (by ≈12.8%) and negative emotion (by ≈20.5%) methods. Overall, as can be seen in Table 1, the RBF kernel results are slightly better than the linear results (≈1.3%).

In this exploratory study, using a small sample size (n=60), receiver operating characteristic curves show a separation classification between the two particular groups of between 86-94%. These models offer an 11-13% improvement over our earlier methods using only emotion or sense categories and supports the idea that an embodied cognition classifier is a better discriminator of the way a person thinks and feels rather than treating the body and mind as separate entities. While the radial based kernel method offers a 1.3% improvement over a linear classifier future research would benefit by not only comparing radical violent individuals with 'normal' bloggers, but also with radical non-violent individuals. These approaches may help classify behavior for influencing online social systems.

ACKNOWLEDGMENTS

REFERENCES

Aloraumy, W., Burnap, P., Liu, H., & Williams, M. (2018). Cyber Hate Classification: 'Othering' Language And Paragraph Embedding. *arXiv preprint arXiv:1801.07495*.

Arras, L., Horn, F., Montavon, G., Müller, K. R., & Samek, W. (2017). "What is relevant in a text document?": An interpretable machine learning approach. *PloS one*, *12*(8), e0181142.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in cognitive sciences*, *7*(2), 84-91.

Bisio, F., Oneto, L., & Cambria, E. (2017). Sentic Computing for Social Network Analysis. In *Sentiment Analysis in Social Networks* (pp. 71-90).

Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, *2*(3), 27.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273-297.

Doyle, C. M., & Lindquist, K. A. (2018). When a word is worth a thousand pictures: Language shapes perceptual memory for emotion. *Journal of Experimental Psychology: General*, *147*(1), 62.

Dreyer, F. R., & Pulvermüller, F. (2018). Abstract semantics in the motor system?–An event-related fMRI study on passive reading of semantic word categories carrying abstract emotional and mental meaning. *Cortex*, *100*, 52-70.

Drodge, E. N., & Murphy, S. A. (2002). Interrogating emotions in police leadership. *Human Resource Development Review*, *1*(4), 420-438.

Egnoto, and Griffin. Analyzing Language in Suicide Notes and Legacy Tokens. *Crisis*, 2016.

Falk, E., & Scholz, C. (2018). Persuasion, influence, and value: Perspectives from communication and social neuroscience. *Annual review of psychology*, *69*.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, *27*(8), 861-874.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics*, *7*(2), 179-188.

Gjelsvik, B., Lovric, D., & Williams, J. M. G. (2018). Embodied cognition and emotional disorders: Embodiment and abstraction in understanding depression. *Journal of Experimental Psychopathology*, *9*(3), pr-035714.

Guell, X., Gabrieli, J. D., & Schmahmann, J. D. (2018). Embodied cognition and the cerebellum: Perspectives from the Dysmetria of Thought and the Universal Cerebellar Transform theories. *Cortex*, *100*, 140-148.

Healey, J. (2014). Physiological sensing of emotion. *The Oxford handbook of affective computing*, 204-216.

Herbert, C., Ethofer, T., Fallgatter, A. J., Walla, P., & Northoff, G. (2018). The Janus Face of Language: Where Are the Emotions in Words and Where Are the Words in Emotions?. *Frontiers in psychology*, *9*.

Huang, H. Y., & Lin, C. J. (2016, June). Linear and kernel classification: When to use which?. In *Proceedings of the 2016 SIAM International Conference on Data Mining* (pp. 216-224). Society for Industrial and Applied Mathematics.

Keerthi, S. S., & Lin, C. J. (2003). Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural computation*, *15*(7), 1667-1689.

Kernot, D., Bossomaier, T., & Bradbury, R. (2017a). *Novel Text Analysis for Investigating Personality: Identifying the Dark Lady in Shakespeare's Sonnets*. Journal of Quantitative Linguistics. Vol 24 No 4, 255-272.

Kernot, D., Bossomaier, T., and Bradbury, R. (2017b). *Identifying Suicide Attackers in Cyberspace.* Presentation at Terrorism and Social Media Conference, Swansea, United Kingdom, 27-29 June 2017.

Kernot, D., Bossomaier, T., & Bradbury, R. (2018). *Shakespeare's Sotto Voce: Determining True Identity from Text* Frontiers in Psychology. Vol 9. March 2018 Article 289, 1-17.

Mahon, B. Z. (2015). What is embodied about cognition?. *Language, cognition and neuroscience*, *30*(4), 420-429.

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences*, 201710966.

Metz, C. E. (1978, October). Basic principles of ROC analysis. In *Seminars in nuclear medicine* (Vol. 8, No. 4, pp. 283-298). Elsevier.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C. C., Lin, C. C. (2017). e1071: Misc Functions of the Department of statistics, Probability Theory Group (Formally E1071), TU Wien. R package version 1.6.8.

Niedenthal, P. M. (2007). Embodying emotion. *science*, *316*(5827), 1002-1005.

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and social psychology review*, *9*(3), 184-211.

Nieuwenhuys, A., & Oudejans, R. R. (2017). Anxiety and performance: perceptual-motor behavior in high-pressure contexts. *Current opinion in Psychology*, *16*, 28-33.

Northoff, G., Heinzel, A., de Greck, M., Bermpohi, F., Dobrowolny, H., Panksepp, J. (2006) Self-referential processing in our brain – a meta-analysis of imaging studies on the self. Neuroimage 2006 May 15;31(1):440-457

Obuchowski, N. A., Lieber, M. L., & Wians, F. H. (2004). ROC curves in clinical chemistry: uses, misuses, and possible solutions. *Clinical chemistry*, *50*(7), 1118-1125.

Pennebaker, J. W., Boyd, R. L., Jordan, K., and Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin. DOI:10.15781/T29G6Z.

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage*, *45*(1), S199-S209.

Picard, R. W., Vyzas, E., & Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE transactions on pattern analysis and machine intelligence*, *23*(10), 1175-1191.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42.

Rani, S., & Singh, J. (2017). Sentiment Analysis of Tweets using Support Vector Machine. International Journal of Computer Science and Mobile Applications, Vol.5 Issue. 10, October- 2017, pg. 83-91

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics*, *12*(1), 77.

Rout, J. K., Choo, K. K. R., Dash, A. K., Bakshi, S., Jena, S. K., & Williams, K. L. (2018). A model for sentiment and emotion analysis of unstructured social media text. *Electronic Commerce Research*, *18*(1), 181-199.

Seyeditabari, A., Tabari, N., & Zadrozny, W. (2018). Emotion Detection in Text: a Review. *arXiv preprint arXiv:1806.00674*.

Singh, J., Singh, G., & Singh, R. (2017). Optimization of sentiment analysis using machine learning classifiers. *Human-centric Computing and Information Sciences*, *7*(1), 32.

Stella, M., Ferrara, E., & De Domenico, M. (2018). Bots sustain and inflate striking opposition in online social systems. *arXiv preprint arXiv:1802.07292*.

Tillman, R., & Louwerse, M. (2018). Estimating Emotions Through Language Statistics and Embodied Cognition. *Journal of psycholinguistic research*, *47*(1), 159-167.

Tzirakis, P., Trigeorgis, G., Nicolaou, M. A., Schuller, B. W., & Zafeiriou, S. (2017). End-to-end multimodal emotion recognition using deep neural networks. *IEEE Journal of Selected Topics in Signal Processing*, *11*(8), 1301-1309.

Varga, S. (2018, March). Embodied concepts and mental health. In *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine* (Vol. 43, No. 2, pp. 241-260). US: Oxford University Press.