INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

ROBERT GORDON UNIVERSITY ABERDEEN

Multimodal Depression Detection and Treatment App

Group 1 Literature Review Document by

Lasal Jayawardena - 20200344
Bhanuka Fernando - 20201006
Lakindu Kumara - 20200987
Mohamed Fenaz - 2019109

Supervised by

Prof. Nihal Kodikara

Submitted in partial fulfilment of the requirements for the BSc(Hons) in Artificial Intelligence And Data Science degree at the Robert Gordon University.

October 2021

Table of Contents

1. Introduction

Within recent years, the percentage of people affected by mood disorders has skyrocketed. The rate of increase in such cases has tripled in the last year alone. Mental health took prominence with the covid-19 pandemic, even though many people faced severe problems before it was in the spotlight. More than 300 million people have been affected globally (World Health Organization, 2017).Until now, there are no centralized mobile or web platforms that could both detect these mood disorders and offer solutions. Our project is introduced to break that stigma. Our project is an integration of several AI components: 1) Depression detection from audio. 2) Semantic-based music recommender system. 3) Sematic-based text classifier 4) The Therapeutic chatbot.

2. Relevant Work

2.1 Depression Detection in Audio

There have been several researches done to identify depression in speech and text. With the Depression Recognition Sub Challenges (DSC) from AVEC2013 to AVEC2018 (Valstar et al., 2016), there was more interest in detecting depression in speech audio as well. Up until now, many techniques were investigated for depression classification and estimation in audio. A binary classification approach was taken by Cohn et al. (2009) where support vector machines (SVM) and logistic regression were used to identify between depressed and non-depressed patients using facial actions and vocal prosody. Another research was conducted by Nicholas et al. (2011), where he tried identifying the key audio features to help detect depression. The paper then showed that an accuracy of above 80% was obtained using the combination of formant-based features and Mel-frequency cepstral coefficient (MFCC).

A deep neural architecture called DepAudioNet was created by Ma et al. (2016), which had the ability to encode characteristics related to depression. Pampouchidou et al. (2016) created a depression classifier that utilized both low level and high level from video, text, and audio modalities. Yang et al. (2017) designed a hybrid Classifier where they used audio, text and video descriptors. Jain et al. (2014) focused on the visual descriptors and utilized a Fisher Vector to estimate the depression levels in the

AVEC2014 (Valstar et al., 2014) where the main task was estimation of the BDI-II scores.

Most of the researches referenced above tried to tackle the classification problem from a single point of view, where it was either classification or estimation of depression. It is important to note that experimental results obtained from (Williamson et al., 2014) (Senoussaoui et al., 2014) (Williamson et al., 2016) (Yang et al., 2016) show that better performance could be obtained when you consider depression estimation and depression classification at the same time. One of the most common datasets used for depression estimation and detection is the DAIC-WOZ Database (Gratch et al., 2014) (Devault et al., 2014).

2.2 Sematic-Based Music Recommendation System

Sentiments/ Emotions is a process of the human brain caused by a person's neurological changes associated with his thoughts, feelings, behavioral responses, and moods. Depression is an emotion that occurs due to despondency. Music psychologists have been regularly examined between music and emotions, and state that music induces a clear emotional response in listeners (Swaminathan and Schellenberg, 2015). In a music recommender system, Lyrics and Audio are the major aspects (Myint and Pwint, 2010) (Laurier, Grivolla and Herrera, 2008), which emotion is focused on, but proved that most of the semantic information is exclusively in lyrics and 29% of people agree with that.

Since the past, humans exclusively use Facial Expressions as a method of Communication or expressing their Emotional state. Sentiment analysis of facial expressions could be induced in music recommendations (Glida et al., 2017). However, results yield in the research on emotion-based music recommendation is not optimal. People tend to share their emotions, opinions, experiences on social media and social networks. According to semantic analysis, people use positive, negative, and neural emotional tags in social networks (Kim, 2013). Based on the emotions/ sentiments extracted via these social networks, music recommendation systems can recommend music based on users' informal texts and sentences posted on social media (Rosa, Rodríguez and Bressan, 2013). As per above, People express their emotions in many methods. But Using Facial Expressions is the most ancient and accurate method while expressing your emotional state in social media becomes popular method today. It would be interesting to analyze how the system recommends music to reduce your depression by using the current appearance of your face and through a successful consideration of your social media posts.

2.3 Sematic-Based Text Classification

Sematic based text classification involves identifying a certain connotation from a corpus of text. There are several key steps when building a semantic based text classifier. First is feature extraction, this is where the documents/text corpora are converted to a structured characteristic space. The most common ways of Feature Extractions are Term Frequency (TF) (Salton and Buckley, 1988), GloVe (Pennington, Socher, and Manning, 2014) and Word2Vec (Goldberg, and Levy, 2014) . Then we need to apply dimensionality reduction that helps efficiently manage memory and inference speeds. Next, we can focus on the classification technique to be used. Finding the best classifier is the most important part of the text classification pipeline. With recent enhancements in GPUs and TPUs, we see a lot of deep neural network architectures used. These deep neural networks are capable to efficiently make more complicated interconnections with data (LeCun, Bengio, and Hinton, 2015). The final component of the pipeline is evaluation. The simplest way of evaluation is the perfection calculation but this will not work for the datasets which are not balanced (Huang, and Ling, 2005).

People often use social media such as Facebook, Instagram and Twitter to express their feeling and opinions and to talk to others. Because of this there are millions of data points being generated which for semantic based text classification. Social media APIs allow the developers to access the data generated on their platform, which is useful to build the training corpus (Coppersmith , Dredze, and Harman, 2014). Support Vector Machines (SVM) is one of the most common machine learning techniques used in semantic text classification, especially for depression prediction. A popular way to generate semantic representation from the text is by using an encoder-decoder architecture called Semantic embedding with Recurrent Neural Networks (Gan et al., 2017). Lately, some phrasebooks are released which can measure the strengths of phrases and words for different emotions. Choudhury (2013) researches a way to predict depression using the social media data, made with tweets of each subject. Pang (2002) used a prior-knowledge-free supervised machine learning method to examine whether semantic text classification was a special case of topic-based categorization.

2.4 Therapeutic Chatbot

Over the years, because of advancements in Machine learning and Artificial intelligence, chatbots are being used as hardware or software for those who seek mental health advice. Several conversational agents are being used to support those who suffer from depression, mental disorders, and anxiety. ELIZA (Weizenbaum, 1966), the first chatbot which examined natural language conversation between man and computer inspired today's chatbots targeting mental health such as Tess, Woebot, Wysa, etc. Abd-alrazaq et al. (2019) reported that outcomes of chatbots and characteristics are inconsistent, which made it difficult to compare the efficiency of chatbots.

According to Russel Fulmer et al. (2018) Tess chatbot was developed by X2 to support people with anxiety and depression using Artificial intelligence. It uses a combination of technologies, emotion algorithms, and machine learning techniques to deliver emotional wellness, although not designed to replace the role of a trained therapist. Wysa is a smartphone-based empathetic artificial intelligence chatbot app that specifically targets symptoms of depressions amongst its users. Becky Inkster et al. (2018) states that Wysa uses evidence-based self-help practices such as CBT, dialectical behavior therapy, motivational interviewing, positive behavior support, behavioral reinforcement, mindfulness, and guided micro-actions and other tools to encourage users to build emotional resilience skills.

Currently, there are only a few Embodied conversational agents (ECAs) capable of early detection and prevention of suicidal behaviors (Martinez, 2017) Help4mood interactive system has been developed with the aim to assist treatment for depression during sessions between patients diagnosed with depression and clinicians, it provides exercises, activities and also implements a crisis plan when suicidal behavior is detected (Burton et al., 2015). As untreated depression is the major cause for suicides, by identifying user's depression through analysis of user's Facebook news feed postings, past text/voice chat history, and web searches, AISA utilize a chatbot model using Seq2Seq and RNN for the emotional wellness of the client to prevent suicidal thoughts and behaviors (Kulasinghe et al., 2019).

3. Comparison Table of Relevant Work

Research	Author	Year	Dataset	Model Used	Metric
Depression Detection in Audio					
Detecting depression from facial actions and vocal prosody	Jeffrey F Cohn, Tomas Simon Kruez, Iain Matthews, Ying Yang, Minh Hoai Nguyen, Margara Tejera Padilla, Feng Zhou, and Fernando De la Torre	2009	Participants (n = 57, 20 men and 37 women, 19% non-Caucasian) from a clinical trial for treatment of depression	SVM	Accuracy: 79%
An Investigation of Depressed Speech Detection: Features and Normalization	Nicholas Cummins, Julien Epps, Michael Breakspear, and Roland	2011	47-speaker depressed/neutral read sentence speech database	GMM	Accuracy: 80%
DepAudioNet: An Efficient Deep Model for Audio based Depression Classification	Xingchen Ma, Hongyu Yang, Qiang Chen, Di Huang, and Yunhong Wang	2016	DAIC-WOZ dataset	CNN, LSTM	F! Score: 0.52(0.70)
Depression Assessment by Fusing High and Low Level Features from Audio, Video, and Text	Anastasia Pampouchidou, Olympia Simantiraki, Amir Fazlollahi, Matthew Pediaditis, Dimitris Manousos, Alexandros Roniotis, Georgios Giannakakis, Fabrice Meriaudeau, Panagiotis Simos, Kostas Marias, et al.	2016	AVEC 2016	Decision Tree	F1 Score: 0.52(0.81)
Hybrid Depression Classification and Estimation from Audio Video and Text Information	Le Yang, Hichem Sahli, Xiaohan Xia, Ercheng Pei, Meshia Cédric Oveneke, and Dongmei Jiang	2017	DAIC-WOZ dataset	Hybrid of DCNN, SVM and Random Forest	F1 Score: 0.667(0.885)
Depression Estimation Using Audiovisual Features and Fisher Vector Encoding	Varun Jain, James L Crowley, Anind K Dey, and Augustin Lux	2014	AVEC 2014	GMM, SVM	MAE: 8.3988
Model Fusion for Multimodal Depression Classification and	Mohammed Senoussaoui, Milton Orlando Sarria	2014	AVEC 2014	SVM, RVM, GLM	Accuracy: 82%

Level Detection. In AVEC@MM.	Paja, João Felipe Santos, and Tiago H. Falk				
Semantic-Based Music Recommendation System					
Smart music player integrating facial emotion recognition and music mood recommendation	Shlok Glida, Hussain Zafar, Chintan Soni and Kshitija Waghurdekar	2017	Images from Kaggle Facial Expression Recognition Challenge	CNN	Accuracy: 90.23%
Music recommendation system based on user's sentiments extracted from social networks	Renata.L.Rosa, Demostenes.Z.Rodriguez and Graca Bressan	2015	User's informal text posts from social networks, 240 songs extracted from Brazilian music portal	SVM	Accuracy: 91%
DEAP: A database for emotion analysis ; using Physiological signals	Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahim, Thierry Pun, Anton Nijholt, Ioannis Patras	2012	Electroencephalogram (EEG) and peripheral physiological signals of 32 participants	RVM	F1 Score: 0.50
Multimodal Music Mood Classification Using Audio and Lyrics	C.Laurier, J.Grivolla and P.Herrera	2008	Database composed of 1000 songs divided between4 categories of interest	SVM, Logistic, Random Forests	Accuracy: 80.7%
Semantic-Based Text Classification					
Quantifying Mental Health Signals in Twitter	Coppersmith G, Dredze M, Harman C	2014	Twitter Data	Linguistic Inquiry Word Count (LIWC)	--
Thumbs up? Sentiment classification using machine learning techniques. Empirical Methods in Natural Language Processing	Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan	2002	Internet Movie Database (IMDb) archive	SVM, Naïve Bayes	Accuracy: 81.6%
Predicting Depression via Social Media	Munmun De Choudhury, M. G. et al.	2013	Twitter Data	MDD classifier	Accuracy: 72.384%

Therapeutic Chatbot					
ELIZA—a computer program for the study of natural language communication between man and machine	Joseph Weizenbaum	1966	Collection of therapist scripts	--	--
Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety	Russell Fulmer, Angela Joerin, Breanna Gentile, Lysanne Lakerink, Michiel Rauws	2018	Data collection from X2AI Inc	--	--
Pilot randomised controlled trial of Help4Mood, an embodied virtual agent-based system to support treatment of depression	Christopher Burton , Aurora Szentagotai Tatar , Brian McKinstry , Colin Mathes`on	2015	21 Help4Mood users	--	Accuracy: 85%
An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study	Becky Inkster, Shubhankar Sarda, Vinod Subramanian	2018	Data collected from users of Wysa	--	Accuracy: 90%

4. Summary

Our project is a Multimodal depression detection and treatment system. We will be using the real-time audio data given by the user and their recent social media history to diagnose depression and determine the current mood of the user. We use techniques of music therapy and psychotherapy to best help each user overcome their emotional problems. We will be building a semantic-based music recommender system and a therapeutic chatbot to integrate the functionalities of our treatment system. Our research gap will mainly focus on detecting depression with audio data. We will experiment with different architectures of parallel structured CNNs and utilize a large variety of mid-level audio features in building our AI model.

5. References

1. *World Health Organization. Depression and Other Common Mental Disorders: Global Health Estimates; Technical Report; World Health Organization: Geneva, Switzerland, 2017.*

2. *M. Valstar, J. Gratch, B. Schuller, F. Ringeval, D. Lalanne, M. Torres Torres, et al., "Avec 2016: Depression mood and emotion recognition workshop and challenge", Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge ser. AVEC '16, pp. 3-10, 2016, [online] Available: http://doi.acm.org/10.1145/2988257.2988258.*

3. *Jeffrey F Cohn, Tomas Simon Kruez, Iain Matthews, Ying Yang, Minh Hoai Nguyen, Margara Tejera Padilla, Feng Zhou, and Fernando De la Torre. 2009. Detecting depression from facial actions and vocal prosody. In Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on. IEEE, 1–7.*

4. *Nicholas Cummins, Julien Epps, Michael Breakspear, and Roland Goecke. 2011. An investigation of depressed speech detection: Features and normalization. In Twelfth Annual Conference of the International Speech Communication Association.*

5. *Xingchen Ma, Hongyu Yang, Qiang Chen, Di Huang, and Yunhong Wang. 2016. DepAudioNet: An Efficient Deep Model for Audio based Depression Classification. In Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge. ACM, 35–42.*

6. *Anastasia Pampouchidou, Olympia Simantiraki, Amir Fazlollahi, Matthew Pediaditis, Dimitris Manousos, Alexandros Roniotis, Georgios Giannakakis, Fabrice Meriaudeau, Panagiotis Simos, Kostas Marias, et al. 2016. Depression Assessment by Fusing High and Low Level Features from Audio, Video, and Text. In Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge. ACM, 27–34*

7. *Le Yang, Hichem Sahli, Xiaohan Xia, Ercheng Pei, Meshia Cédric Oveneke, and Dongmei Jiang. 2017. Hybrid Depression Classification and Estimation from Audio Video and Text Information. In Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge (AVEC '17). Association for Computing Machinery, New York, NY, USA, 45–51. DOI:https://doi.org/10.1145/3133944.3133950*

8. *Varun Jain, James L Crowley, Anind K Dey, and Augustin Lux. 2014. Depression estimation using audiovisual features and fisher vector encoding. In Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge. ACM, 87–91.*

9. *Michel Valstar, Björn Schuller, Kirsty Smith, Timur Almaev, Florian Eyben, Jarek Krajewski, Roddy Cowie, and Maja Pantic. 2014. Avec 2014: 3d dimensional affect and depression recognition challenge. In Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge. ACM, 3–10.*

10. *James R Williamson, Thomas F Quatieri, Brian S Helfer, Gregory Ciccarelli, and Daryush D Mehta. 2014. Vocal and facial biomarkers of depression based on motor incoordination and timing. In Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge. ACM, 65–72.*

11. *Mohammed Senoussaoui, Milton Orlando Sarria Paja, João Felipe Santos, and Tiago H. Falk. 2014. Model Fusion for Multimodal Depression Classification and Level Detection. In AVEC@MM.*

12. *James R. Williamson, Elizabeth Godoy, Miriam Cha, Adrianne Schwarzentruber, Pooya Khorrami, Youngjune Gwon, Hsiang-Tsung Kung, Charlie K. Dagli, and Thomas F. Quatieri. 2016. Detecting Depression using Vocal, Facial and Semantic Communication Cues. In AVEC@ACM Multimedia.*

13. *Le Yang, Dongmei Jiang, Lang He, Ercheng Pei, Meshia Cédric Oveneke, and Hichem Sahli. 2016. Decision Tree Based Depression Classification from Audio Video and Language Information. In Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge. ACM, 89–96.*

14. *Gratch J, Artstein R, Lucas GM, Stratou G, Scherer S, Nazarian A, Wood R, Boberg J, DeVault D, Marsella S, Traum DR. The Distress Analysis Interview Corpus of human and computer interviews. InLREC 2014 May (pp. 3123-3128)*

15. *DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A., Lhommet, M., Lucas, G., Marsella, S., Morbini, F., Nazarian, A., Scherer, S., Stratou, G., Suri, A., Traum, D., Wood, R., Xu, Y., Rizzo, A., and Morency, L.-P. (2014). "SimSensei kiosk: A virtual human interviewer for healthcare decision support". In Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'14), Paris*

16. *Swathi Swaminathan and E. Glenn Schellenberg, "Current emotion research in music psychology", Emotion Review, vol. 7, no. 2, pp. 189-197, Apr. 2015*

17. *E.E.P. Myint and M. Pwint, "An approach for mulit-label music mood classification", 2010 2nd International Conference on Signal Processing Systems, pp. V1-290-V1-294, 2010*

18. *C.Laurier, J.Grivolla and P.Herrera, "Multimodal music mood classification using audio and lyrics", 2008 7th International Conference on Machine Learning and Applications.,2008*

19. *P. N. Juslin and P. Laukka. Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. Journal of New Music Research, 33 (3), 2004.*

20. *S.Glida, H.Zafar, C.Soni and K.Waghurdekar, “Smart music player integrating facial emotion recognition and music mood recommendation”, International Conference on Wireless Communications, Signal Processing and Networking(WiSPNET), 2017*

21. *H. H. Kim, "A semantically enhanced tag-based music recommendation using emotion ontology", Proc. Asian Conference on Intelligent Information and Database Systems, pp. 119-128, Jan. 2013.*

22. *R. L. Rosa, D. Z. Rodríguez and G. Bressan, "Music recommendation system based on user’s sentiments extracted from social networks", IEEE Transactions on Consumer Electronics, Aug. 2013*

23. *Salton, G.; Buckley, C. Term-weighting approaches in automatic text retrieval. Inf. Process. Manag. 1988, 24, 513–523. [CrossRef]*

24. *Pennington, J.; Socher, R.; Manning, C.D. Glove: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; Volume 14, pp. 1532–1543.*

25. *Goldberg, Y.; Levy, O. Word2vec explained: Deriving mikolov et al.’s negative-sampling word-embedding method. arXiv 2014, arXiv:1402.3722.*

26. *LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436–444. [CrossRef] [PubMed]*

27. *Huang, J.; Ling, C.X. Using AUC and accuracy in evaluating learning algorithms. IEEE Trans. Knowl. Data Eng. 2005, 17, 299–310. [CrossRef]*

28. *Coppersmith G, Dredze M, Harman C (2014) Quantifying mental health signals in Twitter. In Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality, 2014, pp. 51–60*

29. *Zhe Gan, Yunchen Pu, Ricardo Henao, Chunyuan Li, Xiaodong He, and Lawrence Carin. 2017. Learning Generic Sentence Representations Using Convolutional Neural Networks. In EMNLP. 2390–2400.*

30. *Munmun De Choudhury, M. G. (2013). Predicting Depression via Social Media. Association for the Advancement of Artificial Intelligence, (p. 10).*

31. *Bo Pang, L. L. (2002). Thumbs up? Sentiment classification using machine learning techniques. Empirical Methods in Natural Language Processing (EMNLP), (p. 7)*

32. *Weizenbaum, J. (1966). 'ELIZA—a computer program for the study of natural language communication between man and machine'. Available: https://dl.acm.org/doi/abs/10.1145/365153.365168*

33. *Abd-alrazaq, A., Alajlani, M., Alalwan. A., and Bewik M. (2019). 'An overview of the features of chatbots in mental health: A scoping review'. Available: https://www.researchgate.net/publication/336054143_An_overview_of_the_features_of_chatbots_in_mental_health_A_scoping_review*

34. *Fulmer, R., Joerin A., Gentile, B., Lakerink, L. and Rauws M. (2018) 'Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety'. Available: https://mental.jmir.org/2018/4/e64/*

35. *Inkster, B., Sarda, S. and Subramanian, V. (2018) 'An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study'. Available: https://mhealth.jmir.org/2018/11/e12106*

36. *Martinez, J. (2017). 'Embodied Conversational Agents for the Detection and Prevention of Suicidal Behaviour: Current Applications and Open Challenges'. Available: https://link.springer.com/article/10.1007/s10916-017-0784-6*

37. *Burton, C., Tatar, A., McKinstry, B. and Matheson, C. (2015). 'Pilot randomised controlled trial of Help4Mood, an embodied virtual agent-based system to support treatment of depression'. Available: https://journals.sagepub.com/doi/abs/10.1177/1357633X15609793*

38. *Kulasinghe, S.A.S.A., Jayasinghe, A., Rathnakaya, A. and Karunarathne, P.B.M.M.D. (2019). 'AI Based Depression and Suicide Prevention System'. Available : https://ieeexplore.ieee.org/document/9103411*