2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) 978-1-7281-4142-8/20/$31.00 ©2020 IEEE 10.1109/ic-ETITE47903.2020.394

Real-time Acoustic based Depression Detection using Machine Learning Techniques

Bhanusree Yalamanchili
Assistant Prof, Dep. of CSE,VNR VJIET
Hyderabad, INDIA
0000-0003-2056-9379

Nikhil Sai Kota
JP Morgan Chase Services India Pvt Ltd
Hyderabad, INDIA
nikhilsai313@gmail.com

Maruthi Saketh Abbaraju
Tata Consultancy Services
Hyderabad, INDIA
amsaket1998@gmail.com

Venkata Sai Sathwik Nadella
Infor
Hyderabad, INDIA
sathwik.nadella@gmail.com

Sandeep Varma Alluri
MS Computer Science, California State University
Fullerton, US
anvsv27799@gmail.com

***Abstract*—Depression disorder is predicted to rise to the second leading cause of disability by 2030 as per the identifications of the World Health Organization (WHO). Though well trained clinicians, medical and psychological treatments are available for depression treatment, persons or families are reluctant to speak out/ reach doctors about this disorder for various social reasons. Diagnosis of depression disorder includes numerous interviews with patient and family, clinical analysis, questionnaires which is time consuming and also demands well trained clinicians. In the present era of Machine learning, automation of depression detection is not complicated and can easily be deployed. However, the automation should use fewer resources, provide accurate results with more reachability. In this paper, acoustic features are used to train a classification model to categorize a human as Depressed or not- Depressed. DIAC-WOZ database available with AVEC2016 challenge is considered for training the classifiers. Prosodic, Spectral and Voice control features are extracted using the COVAREP toolbox and are feature fused. SMOTE analysis is used for overcoming the class imbalance and 93% accuracy is obtained with the SVM algorithm resulting in Depression Classification Model (DCM). An android application cureD Deployed on Cloud is developed to self assess depression using DCM and PHQ-8 questionnaire. The application is tested on real time data of 50 subjects under the supervision of a qualified psychiatrist and an accuracy of 90% is obtained.**

***Index Terms*—Depression detection, Speech processing, SVM, Feature Fusion, Acoustics, SMOTE, DAIC-WOZ , PHQ-8 , Machine Learning**

I. Introduction

Depression is one of the major mental ill health faced by humans of all ages and gender in the recent years. The work culture, stressful life, emotional imbalance,family disturbances, and social life is leading to depression. Depression (Major depression disorder) is becoming common and serious medical illness and is causing negative affects on how one feels and act up on in day to day life. Usually this mental state causes feelings of sadness, loss of interest on things and jobs and rarely may lead to suicide. It affects the natural ability of functioning at work as well as at home. Depression affects one in every 15 adults in a given year and the risk in women is twice than men. The major symptoms includes: changes in food and sleep habits, loss of energy, no concentration, anxiety, hopelessness, feeling of less use, taught of self harming etc. and the major affects include:Gain/loss in weight, heart diseases, inflammation, sexual health problems, chronic health problems, sleep disorders, gastric problems, etc. According to statistics around the world,

> ”322 million people worldwide live with depression”

World health organization (WHO) has identified the unipolar Depression as one of the disability in 2002, and predicted that it may raise to second leading cause by 2030 [1] [2]and has conducted ”Let’s talk campaign” in 2017 to provide awareness on importance of speaking about depression and emotions. Human being a programmed social animal, trained to exhibit only the strengths and misconceptions about being depressed, this mental status is not being openly discussed. Also people confuse between being depressed and sadness/grief/bereavement. Assessing the symptoms and undergoing the treatment immediately is one best solution to overcome the blues of depression. Diagnosis of depression is subjective and depends on openness, support of the individual. Various diagnostic and assessment tools for depression includes, interview style assessment, objective screening mechanism, automatic detection using speech, video and text [3]. Diagnosing and assessment using the first two methods strictly require desire and honesty of a patient as well as highly qualified and trained doctor. The whole process is time consuming and remote diagnosis is also not feasible. The automatic detection from speech, video and text is rather less time consuming and need not require a highly qualified doctor for diagnosing.

The main purpose of the work discussed in the paper is to make the depression detection using speech more feasible in terms of time and resources. The lack of trained doctors can be replaced by trained machine remotely and depression diagnosis can be done effectively. If the individual is detected with depression, then he can be referred to a trained and experienced doctor. In our work the application developed helps in self diagnosing of depression in a very short time with the help of machine learning algorithms. Various acoustic features are extracted using COVAREP toolbox and are fused to enhance the efficiency of classification. Principal Component Analysis (PCA) is used for feature selection.The limitations occurred due to class imbalance in the data set used are over come by SMOTE. Machine learning algorithms like Support Vector Machine, Logistic regression, Random forest and are implemented and their efficiency is cross verified. The model is deployed on to the cloud and the android application is used to collect speech and perform depression detection. Also the diagnosis result is cross verified with analysis result of PHQ-8 questioner. Section 2 focuses on the related work happened in the field of automatic depression detection.Section 3 discuses the universal approaches used to identify depression, while Section 4 deals with the data set used to train the model. Section 5 deals with the Methodology followed for feature engineering, SMOTE and machine learning approaches used. Section 6 depicts the development of Android application and results are provided in section 7.Conclusion and Future scope are discussed in section 8..

II. Related Work

Depression detection approaches have been proposed in many previous works using video, speech and text. Depression Recognition Sub Challenge(DSC) has been released from AVEC2013 to AVEC2018 [4].It is evidently shown that depression of several degrees can be analysed from speech features [5].Correlation between depression and particular features of speech were identified by many researchers [3].

Experiments are performed on acoustic features - prosodic (intensity, pitch, speaking rate, loudness, pause time, jitter, energy etc.), spectral (energy spectrum density, formants, spectral energy distribution, etc.) and cepstral (MFCC and LPCC) [3] to identify their correlation with depression. Glottal features and Teager Energy Operation(TEO) are inspected and found adequate connection with depression identification [3]. Classification accuracy is higher in female speakers when compared to male speakers when TEO and MFCC featuers are considered .Few researchers illustrated the relationship between intensity of depression and speech features [6]. Many researchers has shown that Feature-Fusion and optimization may increase the accuracy of classification [7].
Many researchers tried to improve classification system performance by combining both speech and facial expression of a person which yielded good results [8] [9]. Lang He et al. has proposed Convolution neural networks model for automatic depression detection from speech [10]
. Various approaches for detection of Depression are discussed here briefly.

A. Questionnaire,Interview and Observational Procedures

Detecting depression at primary care set up is a complex and is carried out by a professionally trained psychiatrist. They perform a series of mental tests to confirm the condition and objectively categorize the levels of depression- low, medium or high [3]. Hamilton Rating scale for Depression(HAMD) [11], self-assessment questionnaire like Beck Depression Index(BDI) [12],Patient Health Questionnaire-8 (PHQ-8) [13] are the ones which are universally used by psychiatrists. These methods need willingness and honest response of the individual about his mental situations like interest/pleasure in doing things, Sleeping and eating habits, energy levels, hopelessness etc. The Patient Health Questionnaire-8 (PHQ-8) is an objective self-report questioner containing multiple choices to check mental health disorders of depression based on the score obtained. It is a useful measure for analysing depression state of a person and its diagnostic algorithm with a cut point greater than or equal to 10 can be used for defining current depression [13].Common tools for predicting depression clinically are shown in the Table I

TABLE I
Depression assessment tools

Tool	**Scales**	**Clinical or Self report**	**No.of Questions**
Hamilton Rating Scale for Depression(HAMD)	Normal(0-7), Mild(8-13), Moderate(14-18) Severe(19-22) Very Severe(>= 23)	C	21
Beck Depression Inventory(BDI)	Minimal(0-9), Mild(10-18), Moderate(19-29), Severe(>= 30)	S	21
Patient Health Questionnaire-8(PHQ-8)	Minimal(0-4), Mild(5-9), Moderate(10-14), Moderately severe(15-19), Severe(20-24)	S	8

B. Machine Learning Approaches

Modern trend in detecting depression is using Trained machines by using speech audio, interview videos, images and text responses, etc. [6] . The digital data collected in various forms is analysed and features that are specifically related to depression are extracted. For example, in text responses or transcriptions, words that are frequently used when a person is depressed are identified [8]. In audio data, the specific spectral or cepstral features that signify depression are identified [3]. In video and image data, expressions, eye movement and postures are identified [8]. Various machine learning algorithms like K Nearest neighbors(KNN), Gaussian Mixture Model(GMM), Support Vector Machine(SVM) are widely used to investigate depression using different speech types and emotions [6]. Neural networks are used to train machines for early detection of depression using linguistic and text sequences [14].

III. Speech Corpus

A. Distress Analysis Interview Corpus-Wizard-of-Oz Interviews(DAIC-WOZ)

Dataset related to depression audio samples are taken from DAIC-WOZ. This database is a part of larger corpus and stands for “Distress Analysis Interview Corpus (DAIC)”. It contains clinical interviews that help in the diagnosis of distress conditions such as depression, anxiety and post-traumatic stress disorder . These interviews are performed in order to create a intelligent computer agent that is capable of interviewing people and identifying their “verbal and nonverbal indicators of mental illness”. The released data corpus includes audio and video recordings,trancripts of responses. Many verbal and non-verbal features are available in the data provided as a part of interviews. The data includes 189 sessions of interactions ranging between 7 and 33 minutes (average of 16 minutes). Each session includes transcripts of the interaction, participant audio files, and facial features.

B. AVEC2016 – Real Life Depression Challenge

As a part of Audio-Visual Emotional Challenge and Workshop-2016 (AVEC2016) Depression Classification Sub-Challenge(DSC) is released. Along with the challenge, Depression analysis corpus, and baseline results using random forest regression are provided.The F1 score, precision and recall for depressed and not depressed classes on test set [4] are given in the Table II. Though the results show that video alone is performing well, our work focuses on improving the performance on Audio using various machine learning algorithms. Choosing video has its own drawbacks of ethical issues, high storage access when working remotely.

TABLE II
Base line results of Depression Classification Sub-Challenge(DSC) in AVEC2016. Values for class depressed are given in brackets.

Modality	F1	Precision	Recall
Audio	0.582(0.410)	0.941(0.267)	0.421(0.889)
Video	0.851(0.583)	0.938(0.467)	0.790(0.778)
Ensemble	0.857(0.583)	0.938(0.467)	0.790(0.778)

IV. Methodology

A. Feature Engineering

Feature engineering in speech includes proper extraction of features frame wise, calculating the statistical measures, feature fusion and their selection based on correlation. 73 baseline Low Level Descriptors(LLD) including Prosodic(2), Voice quality(8) and Spectral(63) features [15] are extracted using COVAREP toolbox [16] for each frame sampled at 10ms interval. Statistical measures- mean, median, variance, kurtosis, skewness, minimum, maximum, standard deviation are calculated on all the features resulting in 584 feature set. The features extracted and statistical functions performed are summarized in Table III. Prosodic, Spectral and Voice Quality features are fused [17] and Principal Component Analysis (PCA) is applied to reduce feature space dimension.

TABLE III
LLD’s extracted from Speech Corpus and Statistical functions

Low- Level Descriptors(73)
F0,Normalized amplitude quotient-NAQ, Quasi Open Quotient-QOQ, Differentiated Glottal Source Spectrum- H1H2, Parabolic spectral parameter-PSP, Maxima dispersion quotient-MDQ, Peak slope, Glottal pulse dynamics- Rd,Mel Cepstral coefficients- MCEP 0-24, Harmonic model and Phase distortion mean -HMPDM 0-24, Harmonic model and Phase distortion Deviations- HMPDD 0-12
Statistical functions(8)
mean, median, variance, kurtosis, skewness, minimum, maximum, standard deviation

B. Model Development

DIAC-WOZ corpus has 189 subjects among which, 133 subjects are non depressed and 56 are depressed. The class imbalance, due to less availability of depressed subjects can be observed in this corpus. The data set is split in to 60%, 15% and 25% partitions training, development and testing respectively. The fused feature set of the training portion is used to train models using Support Vector Machine(SVM), Logistic Regression and Random Forest. The class imbalance in the data set has generated biased results that are shown in Table IV. To overcome this problem Synthetic Minority Oversampling Technique(SMOTE) is applied. Accuracy obtained for various classifiers with and without SMOTE analysis are shown in Table V. The framework used to design the classification model for Depression detection is shown in Fig. 1.

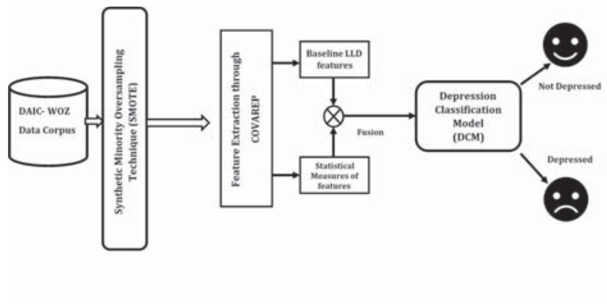

Fig. 1. Framework for generating Depression Classification Model(DCM).

C. Synthetic Minority Oversampling

Any classification algorithm with class-imbalance in data, results in biased prediction in favour of the majority class. In higher dimensional data, as the number of variables greatly exceeds the number of samples, the bias towards majority class is even larger. Performing under sampling of majority class or oversampling minority class, results in class-balanced data.“Synthetic Minority Oversampling Technique (SMOTE)” is an oversampling method, that achieves random oversampling to balance the classes in data [18]. Yet, thorough study of its behaviour on high-dimensional data is to be further explored.

In our work prior to the application of SMOTE analysis, the training data had 132 non-depressed and 56 depressed samples. After applying SMOTE analysis and oversampling the minority class (depressed class), the training data now contains 132 non-depressed and 132 depressed samples.Training, development and testing sets Partitions ratio is remained same as before SMOTE application.After SMOTE analysis is done, number of depressed class observations are balanced with the number of not depressed class observations. All the classification algorithms are performed again on SMOTE applied data and improvement in accuracy and other performance metrics is observed.

TABLE IV
PERFORMANCE METRICS OF CLASSIFIERS ON CLASS IMBALANCE DATA SET AND SMOTE APPLIED DATA SET.VALUES FOR CLASS DEPRESSED ARE GIVEN IN BRACKETS

Classifier	**With class imbalnce**			**With SMOTE analysis**		
	Precision	**Recall**	**F1**	**Precision**	**Recall**	**F1**
Logistic Regressi.	0.77 (0.41)	0.70 (0.50)	0.73 (0.45)	0.79 (0.42)	0.67 (0.57)	0.72 (0.48)
Random Forest	0.67 (0.69)	0.6 (0.7)	0.52 (0.51)	0.47 (0.53)	0.62 (0.71)	0.54 (0.41)
SVM .	0.79 (0.63)	0.70 (0.42)	0.88 (0.59)	0.95 (0.70)	0.93 (0.78)	0.94 (0.74)

TABLE V
ACCURACY OF CLASSIFIERS ON CLASS IMBALANCE DATA SET AND SMOTE APPLIED DATA SET.

Classifier	**Accuracy in % with class imbalance**	**Accuracy in % after SMOTE analysis**
Logistic Regression	63.8	65
Random Forest	59.7	61.7
SVM	70.2	93

D. Logistic Regression

Usually logistic regression is used on data when the dependent variable is categorical in nature. Since the class label in our data is binary (depressed or not depressed), Binary logistic regression is used. For implementing the classifier newton-cg is used as solver with multinomial class and 100 iterations.

E. Random Forest

Random forest is an ensemble learning technique used to create classification models by constructing number of decision trees while training, and returns the mode of the classes of the individual decision trees constructed. The parameters we used in the random forest model are gini as criterion and 10 estimators.Though the performance is comparatively lower than other classifiers, it is implemented to compare with the baseline results of AVEC-2016 challenge.

F. Support Vector Machine

Support Vector Machine(SVM) algorithm tries to find a best hyperplane that separates classes. SVM uses kernels which are transformation functions that convert the data into higher dimensionality to find a hyperplane that can classify the data and this classification might not be possible in lower dimensions. The new data given to classifier is also transformed to higher dimensions using the kernel and then it is plugged in the hyperplane equation to identify the class it belongs to [19]. RBF Kernel is used with gamma scale in this work.

All these three classification techniques namely Logistic Regression, Random Forest and SVM are implemented in SCIKIT-LEARN toolbox. The evaluation metrics are calculated individually for each classifier- over speech of DIAC-WOZ data corpus and are shown in Fig. 2. Finally "Depression Classification Model(DCM)" is generated with an accuracy of 93% on validation data set using acoustic features.

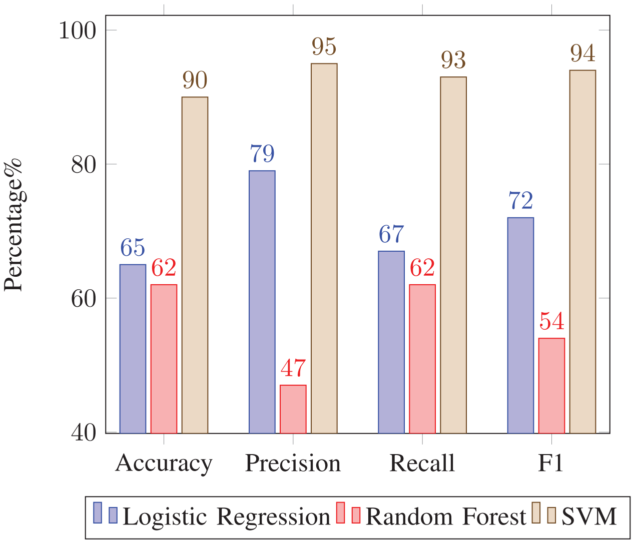

Fig. 2. Depression classification results for the data set after SMOTE Analysis.

V. ANDROID APPLICATION

An user interface (android app) "cureD" is created so that any user can utilise DCM and self-assess their state of depression. The general inertia in talking about one's mental state and reluctance in approaching doctors for clinical analysis can be over come by this application. By using this application even remote assessment of depression can be performed by a psychiatrist where his physical presence is not possible. Any register user by downloading the android application can utilise the Depression Detection Application -"cureD" and assess their state of depression. Even the PHQ-8 result obtained from cureD may help the doctor to evaluate the depression level(mild, moderate, severe etc.)of the positive cases.

User can register and use the application whereas doctors will have a separate login and doctor's role here, is to answer the queries posted by the users and evaluate the PHQ-8 results in case of necessity.

The register user will first answer a standard PHQ-8 questionnaire, followed by recording the voice by reading out the passage displayed on the screen. The person's speech and responses of the questionnaire are uploaded to the server. The trained Depression Classification Model(DCM)which uses SVM algorithm is deployed on WAMP server on Amazon Web Service(AWS) cloud. The Person's speech is preprocessed to remove any noise present during the recording process.73 features are extracted using COVAREP toolbox and their statistical measurements are also fused. The resultant feature vector is used to test the persons depression state.The DCM now evaluates the feature spectrum and classification result is generated.The PHQ-8 score is also calculated in the server. The results of DCM and PHQ-8 are fetched from the server to the mobile application and are displayed to the person. In case the person requires assistance he can post to the doctor on cureD. The doctor can know verify the results and assist the individual for further treatment.The flow diagram for cureD is shown in Fig. 3

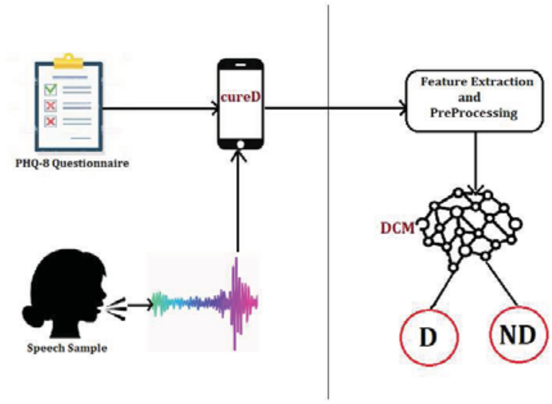

Fig. 3. Flow of cureD Application.

VI. Results

For testing the model in real-time, data is collected from 50 different individuals with ethical considerations. They were asked to fill out the PHQ-8 questionnaire and give their voice sample through the android application cureD mentioned above. All these 50 individuals are examined and classified by a professional psychiatrist with the help of clinical assessment tools, interviews and physical examination. The classification labels were also examined using their PHQ-8 scores, where scores greater than or equal to 10 were labelled as depressed and scores less than 10 were labelled as not depressed. Out of 50 persons, 41 people were labelled as not depressed and 9 persons were labelled as depressed based on their PHQ-8 scores as well as by professional psychiatrist.

The trained DCM model is made available on AWS cloud and is released as an android application. This mobile application is fetched with the speech signals of these 50 individuals and are tested. The model has classified the subjects and the confusion matrix generated is shown in Table VI. This trained model DCM has given 90 % of accuracy and the results are given in Table VII. ROC (Receiver Opersating Charecteristocs) curve to show the relation between the true positive rate and the false positive rate is shown in Fig. 4.

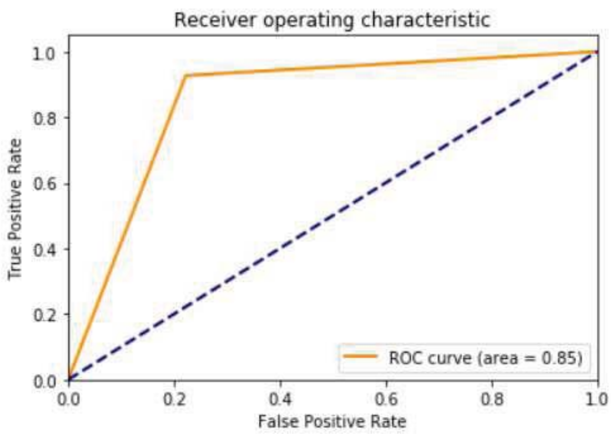

Fig. 4. ROC Curve for SVM.

TABLE VI
Confusion matrix of DCM on Real-time data

	Predicted 0	*Predicted 1*
Expected 0	38	3
Expected 1	2	7

TABLE VII
Evaluation metrics of DCM

Class	*Precision*	*Recall*	*f1-score*	*Accuracy*	*RMSE*
0	0.95	0.93	0.94	90%	0.31
1	0.70	0.78	0.74		

VII. Conclusion and Future Works

The 'DAIC-WOZ' data set considered for this work (available at AVEC 2016 challenge) has class imbalances. The class imbalance is due to the less availability of 'depressed' audio samples in comparison with 'not depressed' audio samples. This leads to biased machine learning model. In order to overcome the class imbalance problem, SMOTE analysis is used.The results of various classification algorithms like Logistic Regression, Random Forest and SVM are compared. Among all the other classifiers SVM after SMOTE analysis has provided good performance.

The conventional methods followed by psychiatrists in detecting depression includes clinical examination,interviews, interaction with family members, responses for various questionnaires like HAMD, BDI, PHQ-8 etc. Due to several social issues persons with depression or their family members hesitate to approach doctor for treatment. This delay may cause severe problems at home and work place and worsen the human relations, which may further result in severe depression. Better treatment can be provided if diagnosis happen faster. To enable this an android application cureD for self assessment along with doctor assistance is created. The application is provided on cloud to enable remote access to patient as well as doctor. The trained model DCM using SVM algorithm is used for Real Time Depression Diagnosis with an efficiency of 90 %. The main advantage of the application is that users need not visit a doctor unless the results turn up positive for depression.

The Dataset which is considered for this work is completely based on foreign accent and the model is trained to it . The data set could be extended to the various native accents and languages by adding audio samples to the existing data set. Deep Neural networks can be implemented to generate a model to classify depression, which may result in more accuracy and adopt to various accents,languages, gender and age.

References

[1] W. H. Organization *et al.*, “The global burden of disease: 2004 update,” 2008.

[2] C. D. Mathers and D. Loncar, “Projections of global mortality and burden of disease from 2002 to 2030,” *PLoS medicine*, vol. 3, no. 11, p. e442, 2006.

[3] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, and T. F. Quatieri, “A review of depression and suicide risk assessment using speech analysis,” *Speech Communication*, vol. 71, pp. 10–49, 2015.

[4] M. Valstar, J. Gratch, B. Schuller, F. Ringeval, D. Lalanne, M. Torres Torres, S. Scherer, G. Stratou, R. Cowie, and M. Pantic, “Avec 2016: Depression, mood, and emotion recognition workshop and challenge,” in *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*, ser. AVEC ’16. New York, NY, USA: ACM, 2016, pp. 3–10. [Online]. Available: http://doi.acm.org/10.1145/2988257.2988258

[5] N. Cummins, J. Epps, M. Breakspear, and R. Goecke, “An investigation of depressed speech detection: Features and normalization,” in *Twelfth Annual Conference of the International Speech Communication Association*, 2011.

[6] H. Jiang, B. Hu, Z. Liu, L. Yan, T. Wang, F. Liu, H. Kang, and X. Li, “Investigation of different speech types and emotions for detecting depression using different classifiers,” *Speech Communication*, vol. 90, pp. 39–46, 2017.

[7] S. Kuchibhotla, H. Vankayalapati, R. Vaddi, and K. R. Anne, “A comparative analysis of classifiers in emotion recognition through acoustic features,” *International Journal of Speech Technology*, vol. 17, no. 4, pp. 401–408, 2014.

[8] A. Haque, M. Guo, A. S. Miner, and L. Fei-Fei, “Measuring depression symptom severity from spoken language and 3d facial expressions,” *arXiv preprint arXiv:1811.08592*, 2018.

[9] J. R. Williamson, D. Young, A. A. Nierenberg, J. Niemi, B. S. Helfer, and T. F. Quatieri, “Tracking depression severity from audio and video based on speech articulatory coordination,” *Computer Speech & Language*, vol. 55, pp. 40–56, 2019.

[10] L. He and C. Cao, “Automated depression analysis using convolutional neural networks from speech,” *Journal of biomedical informatics*, vol. 83, pp. 103–111, 2018.

[11] M. Hamilton, “The hamilton rating scale for depression,” in *Assessment of depression*. Springer, 1986, pp. 143–152.

[12] W. Katon and H. Schulberg, “Epidemiology of depression in primary care,” *General hospital psychiatry*, vol. 14, no. 4, pp. 237–247, 1992.

[13] K. Kroenke, T. W. Strine, R. L. Spitzer, J. B. Williams, J. T. Berry, and A. H. Mokdad, “The phq-8 as a measure of current depression in the general population,” *Journal of affective disorders*, vol. 114, no. 1-3, pp. 163–173, 2009.

[14] M. Trotzek, S. Koitka, and C. M. Friedrich, “Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences,” *IEEE Transactions on Knowledge and Data Engineering*, 2018.

[15] F. Eyben, A. Batliner, and B. Schuller, “Towards a standard set of acoustic features for the processing of emotion in speech.” in *Proceedings of Meetings on Acoustics 159ASA*, vol. 9, no. 1. ASA, 2010, p. 060006.

[16] G. Degottex, J. Kane, T. Drugman, T. Raitio, and S. Scherer, “Covarep—a collaborative voice analysis repository for speech technologies,” in *2014 ieee international conference on acoustics, speech and signal processing (icassp)*. IEEE, 2014, pp. 960–964.

[17] K. R. Anne, S. Kuchibhotla, and H. D. Vankayalapati, “Feature fusion techniques,” in *Acoustic Modeling for Emotion Recognition*. Springer, 2015, pp. 27–32.

[18] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote: synthetic minority over-sampling technique,” *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[19] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2001.