A Chronic Psychiatric Disorder Detection Using Ensemble Classification

V. J. Jithin, G. Manoj Reddy, R. Anand, and S. Lalitha(✉)

Department of Electronics and Communication Engineering,
Amrita School of Engineering, Bengaluru, Amrita Vishwa Vidyapeetham,
Bengaluru, India
jithinvj06@gmail.com, gmanojreddy1996@gmail.com, anand.ramandran@gmail.com,
s_lalitha@blr.amrita.edu

Abstract. The objective of this work is to detect depression that is a more chronic psychiatric disorder found in humans using speech samples. This work is first of its kind in depression detection using Audio Visual and Emotional Challenge 2011 (AVEC 2011) and Computational and Paralinguistics challenge 2016 (ComParE 2016) feature sets. A novel method of ensemble classification using simple machine learning algorithms of Instance-Based classifier with parameter K (IBK), Stochastic Gradient Descent (SGD) and Random Forest is proposed for the projected task with gender dependent and independent systems. Experimental results demonstrate the superiority of ComParE 2016 over AVEC 2011 in determining the psychological state of an individual. Feature selection method is applied to reduce feature vector size, maintaining the accuracy of depression detection with that obtained using large size feature sets. The ensemble-based classification provide better accuracy performance than the individual classifiers.

Keywords: AVEC 2011 · ComParE 2016 · Ensemble voting classifier · IBK · Random Forest · SGD

1 Introduction

World Health Organization discerned that depression is a psychiatric aberration that leads to low spirit, loss of pleasure or interest, low energy levels, poor concentration and bad appetite [1]. Anxiety and depressive disorders are highly prevalent around the world. Different ways of depression detection are proposed in the past time. Many of the depression detection technique uses psychological and behavioral signals for characterizing depression. These include facial expression, body gestures, eye movement, speech, etc. [2]. Amongst these, speech is considered to be an effective method to detect depression since it is simple and non-invasive. Conventional method of depression detection using speech involve feature extraction followed by classification. The prosodic and acoustic features of the speech signals are extracted and fed into various classifiers viz Support

S. M. Thampi et al. (Eds.): SIRS 2019, CCIS 1209, pp. 173–185, 2020.
https://doi.org/10.1007/978-981-15-4828-4_15

Vector Machines (SVM), Hierarchical Fuzzy Signature (HFS), Gaussian Mixture Models (GMM), Multilayer Perceptron (MLP), etc. [3]. These classifiers predict the speech samples as depressed or non-depressed. Literature reveals various attempts by research community in depression detection. Moore et al. achieved a classification accuracy from 87% to 94% using glottal waveform features for differentiating a patient group and control group [4]. Ooi used a major voting classifier on a parallel ensemble. Depression in adolescents was predicted using a multichannel weighted classifier, giving an accuracy of 73% [5]. Teager Energy Operator (TEO) and glottal features provide an accuracy of 69% for predicting major depression in adolescents in the work proposed by Ooi [6]. Ozdas et al. distinguished participants as suicidal, major depressed and healthy by investigating their glottal flow spectrum and vocal jitter using spectral slope getting an accuracy of 90% [7]. Alghowinem used different classifiers: Support Vector Machines (SVM), Hierarchical Fuzzy Signature (HFS), Gaussian Mixture Models (GMM), and Multilayer Perceptron (MLP) and found a hybrid combination of SVM and GMM giving better results [3]. In a proposed work by Yingthawornsuk [8], two speaking styles were used: a text-reading in and a clinical interview. On analyzing, it was found that depressed subjects possessed a slow rate of speech, monotonous delivery and a low range of fundamental frequency. Hence, they formulated small changes in characteristics of speech (e.g. rate of speaking, loudness, differences in pitch etc.) as indicators. Stress is one of the major factor that leads to psychological disorder. Narayanan et al. in their work [9] used the Gammatone frequency cepstral coefficients and multiclass SVM for stress detection. The non-linear TEO features classified using Probabilistic Neural Network (PNN) and the Multi-Layer Perceptron Neural Network (MLPNN) also achieved better performances. Lalitha et al. in [10] used Mel and Bark scale dependent perceptual features to detect emotion. The proposed model outperformed other models in terms of recognition accuracy. K Nearest Neighbor (KNN) and Simple Logistic Classifier (SL) provided an accuracy of 84.7% for affective state recognition in the work proposed by Murali et al. [11]. The Mel Frequency Cepstral Coefficients (MFCC), Perceptual Linear Predictive Cepstrum (PLPC), and Mel Frequency Perceptual Linear Prediction Cepstrum (MFPLPC) are found to play a prominent role for SER as proposed by Lalitha [12]. From the survey, it is evident that the performance of the individual classifiers proved to provide less accuracy performance comparatively. It is observed that there are setbacks in advanced warning techniques to acknowledge the depressed person. Due to less combinations and ensemble pruning methods, the performance of the systems are less efficient. This resulted in less accuracy performance between 60% to 70%. The ComParE feature set have been used in the automatic recognition of eating conditions in speech, that is whether the people are eating while speaking. The average recalls obtained using ComParE features witnessed above 90% [13]. The AVEC features set have been used in building artificial listener agents that can captivate a person in emotionally colored conversation [14]. This work contributes a method to detect the psychiatric disorder of an individual. The focus of this work is on investigation of feature sets i.e ComParE 2016 over

AVEC 2011 and classifier combinations for the detection of psychiatric disorder. A search to find the appropriate feature set and classifier is proposed here. The work also focuses on comparing the performance of the individual classifiers like IBK, SGD, Random Forest and an ensemble of these classifiers using the appropriate feature sets in determining the psychiatric state.

This work is organized in five sections. Section 2, explains the proposed work for depression detection and performance metrics used. In Sect. 3, experiment and analysis is given. Section 4 has a comparative analysis. Finally, Sect. 5 discusses the conclusion and future scope.

2 Proposed Work for Depression Detection Using Ensemble Classification Method

The proposed model for depression detection using ensemble classification consists of five stages - speech database, pre-processing, feature extraction, feature selection and classification. The flow chart of the proposed model is depicted in Fig. 1.

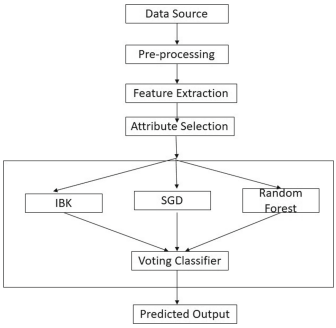

Fig. 1. Flow chart of proposed work

2.1 Speech Database

The database used in this work is DAIC-WOZ depression database. This database includes questionnaire responses and audio recordings, incorporated from a larger collection, the Distress Analysis Interview Corpus (DAIC). It comprises of 189 sessions of interactions of duration 7 to 33 min each. Speech samples were recorded from 53 speakers, of which 31 male and 22 female speakers were used [15].

2.2 Pre-processing

The audio of the interviewer and interviewee are separated initially. Audio file corresponding to each speaker (interviewee) are divided into segments of 10 s. This is performed with help of an audio tool Audacity [16]. A total of 1763 audio sub-samples were segmented from the samples of 53 interviewees. Features for each segment are extracted individually. Waveform of the depressed and non-depressed interviewee is depicted in Figs. 2 and 3 respectively. It can be observed that non-depressed voice sample has no large variations in amplitude levels while the depressed voice sample has variations in intensity by a large extent.

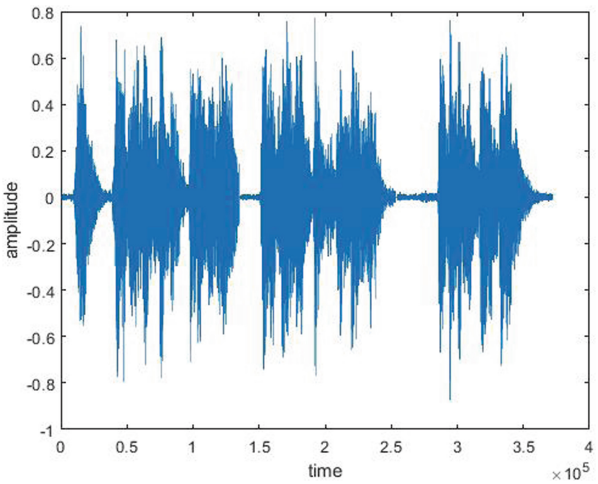

Fig. 2. Non depressed speech sample

2.3 Feature Extraction

The ComParE 2016 [17] and AVEC 2011 [18] features are extracted from the audio samples using openSMILE [19]. openSMILE is an open source software for automatic extraction of features. The software is likewise fit for perceiving the attributes of a given speech or music fragment. The Computational and Paralinguistics challengE 2016 (ComParE 2016) feature set consists of 6373 features. Energy, spectral, cepstral (mfcc), and voice related low level descriptors are few among those. These features come mainly under three groups. First group comprises of 4 energy LLDs, 55 spectral LLDs and its delta coefficients. Second group consists of 6 voice related LLDs and delta coefficients. A lot of 54 functionals is connected to LLDs of first group and 46 functionals to its delta coefficients,

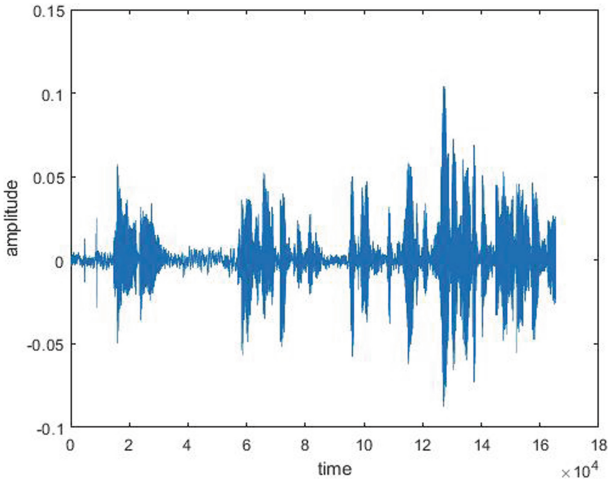

Fig. 3. Depressed speech sample

bringing about $59^*(54+46)=5900$ features. A lot of 39 functionals is connected to second group and its delta coefficients, bringing about $6^*(39+39)=468$ features. The third group comprises of five static descriptors, registered for voiced portions. Along these lines the final set comprises of $5900+468+5=6373$ features [13]. An insight of the ComParE 2016 feature set is mentioned in Table 1.

Table 1. Insight of ComParE 2016 features

Sl. no.	Group	Functionals on LLDs	Functionals on delta coefficients	ComParE features
1	4(energy) + 54(spectral) = 59 LLDs	54	46	5900
2	6(voice related) LLDs	39	39	468
3	5(temporal static descriptors)	–	–	5
4	Total Features			6373

The Audio Visual and Emotional Challenge 2011 (AVEC 2011) feature set comprises of 1941 highlights. It incorporates 25 spectral and energy related low-level descriptors, 6 voicing related LLDs, 25-delta energy/spectral LLD, 6 delta coefficients of the voice related LLDs and 10 voiced/unvoiced durational highlights [14]. An insight of AVEC 2011 feature set is mentioned in Table 2. These features are extracted for the entire 1763 audio samples and further fed into the classifiers.

Further the feature selection method is employed to select the relevant features.

Table 2. Insight of AVEC 2011 features

Sl. no.	Group	Functionals on LLDs	Functionals on delta coefficients	AVEC 2011 features
1	25 LLDs(energy and spectral)	42	23	1625
2	6(voicing related) LLDs	32	19	306
3	10(voiced/unvoiced durational features)	–	–	10
4	Total Features			1941

2.4 Feature Selection

Among the features obtained after feature extraction, some features contribute more to the prediction output than certain other features. These features needed to be selected and this helps reduce the complexity of the classification. Thus to select those relevant features, attribute selection is used. The feature selection is carried out using WEKA [20] in two parts, Attribute Evaluator and Search method. In this work, cfssubsetEval Attribute evaluator along with Best First Search method is used. cfssubsetEval is a correlation based feature selection subset evaluator. Different combinations of attributes are looked into and the best attributes are chosen. The Best First [21] search technique looks through the space of attribute subsets by slope climbing enlarged with a backtracking facility. Setting the quantity of continuous non-improving hubs controls the dimension of backtracking done. Best First strategy begins with either an unfilled set of attributes and inquiry forward or begin with full set of attributes and hunt in reverse, or begin anytime and look in both directions. After performing the attribute selection, classification of the samples into depressed and non-depressed are performed.

2.5 Classification

Classification is a significant data mining strategy with wide applications. It is used to organize everything in a set of data into one of predefined set of classes or groups. The classification is performed using SGD, IBK, Random Forest and an ensemble voting classifier in Weka.

SGD utilizes stochastic slope drop to learn models. Stochastic Gradient Descent is a steady calculation that can be connected to bigger datasets. It replaces every single missing values and transforms nominal attributes into binary ones. It replaces each and every missing values and changes nominal features into binary ones. It additionally standardizes all characteristics, so the coefficients in the yield depend on the standardized information. This ensures better prediction [22].

IBK is a case based classifier which finds the preparation case nearest in the Euclidean separation to the given test tests and predicts the comparable class as this preparation occurrence. It's a k-closest neighbor classifier which decides the quantity of closest neighbors naturally utilizing leave-one-out cross-validation or can be indicated in the object editor. A kind of different search algorithms can be utilized to quicken the assignment of finding the closest neighbors. A linear search is the default yet further choices fuse search trees, for example, KD-trees and ball trees. Predictions from more than one neighbor can be weighted by their distance from the test occasion and after that changing over the distance into a weight [23].

Random Forest is a combination of tree predictors and an ensemble learning method for classification. A multitude of decision trees are used for the classification. Each individual tree performs the class prediction from which the final class of the Random Forest classifier is predicted using a voting classifier of the individual classifiers. Random Forest provides a powerful and effective prediction solution for many practical solutions [24].

Ensemble Voting Classifier is a meta classifier for combining different classifiers for classification via majority voting. Each of the classifiers selected will predict a class label for a test sample. The label which is predicted the most will be selected as output of the voting classifier. An outfit of SGD, IBK, and Random Forest have been ensembled to build up the proposed model.

2.6 Performance Metric

In this work, performance of the depression detection system is evaluated by determining accuracy. The weighted average of recall of the depressed and non-depressed classes gives the accuracy of the system.

$$Accuracy = \frac{TruePositive}{FalseNegative + TruePositive}$$

True positives are information focuses delegated positive by the model that really are positive (which means they are correct), and false negatives are information focuses the model distinguishes as negative that are really positive (off base) [25].

3 Experiment and Analysis

Experimentation is performed to analyze the performance analysis of proposed method using ensemble classification for depression detection on the DAIC-WOZ speech dataset. Speech feature sets derived from ComParE 2016 and AVEC 2011 are referred to as feature set 1 and feature set 2 respectively. ComParE 2016 feature set consists of 6373 features and AVEC 2011 feature set consists of 1941 features respectively. SGD, IBK and Random Forest classifier will comprise the individual classifiers while combination of all these three result in ensemble classifier.

3.1 Proposed Approach for Gender Dependent System (Experiment 1)

Comprehensive test was performed on male and female speakers independently to analyze the performance of ensemble classifier. Two data sets from DAIC-WOZ database containing 1084 voice samples from 31 male speakers and 703 voice samples from 22 female speakers respectively were tested independently with ensemble and individual classifiers for both feature sets. Comparison of accuracy performance for male and female speakers are shown in Tables 3 and 4 respectively.

Table 3. Accuracy (%) analysis for male speakers using feature-set 1 and feature set 2

Classifier	Accuracy using feature set 1	Accuracy using feature set 2
SGD	92.4	73.6
IBK	91	79.1
Random Forest	91.4	79.9
Ensemble classifier	95.6	80.6

Table 4. Accuracy (%) analysis for female speakers using feature-set 1 and feature set 2

Classifier	Accuracy using feature set 1	Accuracy using feature set 2
SGD	95.4	82.7
IBK	92.9	85
Random Forest	92.2	84.1
Ensemble classifier	96.0	85.4

From Table 3 for male speakers, depression is well detected using ensemble method with a 3% rise from the result provided by the individual classification using feature set 1 from SGD. Table 4 depicts with female speakers an improvement of 0.6% is observed with ensemble classifier against SGD using feature set 1. Thus, ensemble classifier provided best results for both male and female speakers. As for gender dissimilarity, the classification accuracies for female speakers for both feature sets were higher than that of the male speakers since females are more expressive. The next experiment involved to analyze the proposed method on depression detection using gender independent system.

3.2 Proposed Approach for Gender Independent System (Experiment 2)

In the experiment 2, the accuracy analysis of the system for entire data set using both the feature sets is compared in Table 5. The system performance using

feature set 1 is considerably higher and error is narrowed to 10% for individual classifiers and to a 5% margin for ensemble classifier. So, it is evident from experiments 1 and 2 the features from the ComParE 2016 feature set provides much better assistance while predicting the depression state of the person than that of AVEC 2011 feature set. So, in further experiments classification will be performed only using feature set 1. The narrow deprivation in the performance of ensemble classifier when compared to the results of experiment 1 conveys that system performance is more effective when classification is carried out for different genders separately.

Table 5. Accuracy (%) analysis for different classifiers using feature set 1 and feature set 2.

Classifier	Accuracy using feature set 1	Accuracy using feature set 2
SGD	91.3	71.0
IBK	89.4	78
Random Forest	91.7	76.8
Ensemble classifier	94.3	78.5

Also, in both the experiments it is observed that feature set 1 that constitute ComParE 2016 features are more effective for depression detection using speech. Henceforth, the work is carried on using feature set 1. Owing to the huge size of feature set 1, the following experiments involve application of feature selection and analyze the proposed system performance on gender dependent and gender independent context.

3.3 Proposed Approach for Gender Dependent and Independent System Using Feature Selection (Experiment 3)

In experiment 3, attribute selection classifier was used that employs the strategy of best search method to find the most relevant features from 6373 features of the feature set 1 for gender dependent scenario with male and female speakers as well as gender independent case. It was found that 130 features were selected. These features responded better with depression state. Feature selection was imparted on data sets for male and female separately. The accuracy obtained by the system with and without feature selection are interpreted in the Table 6. It can be observed that the ensemble method dominated the best individual classifier results around 3% and 0.6% for male and female speakers respectively while approximately 2% rise achieved with gender independent samples. That though there is a narrow deprivation in systems performance using feature selection, significant performance was achieved using compact feature set. Using feature

selection systems performance for female speakers is slightly higher than male speakers. With feature selection application, although around 1% reduction in accuracy is observed against the results of experiment 1 of Tables 3 and 4 of gender dependence while 1.5% accuracy dip attained with gender independence of Table 5, but size of the feature vector is drastically reduced from 6373 to 130 features.

Table 6. Accuracy (%) analysis for different classifiers using feature set 1 using feature selection.

Classifier	Male speakers	Female speakers	Gender independent
SGD	91.4	94.1	89.5
IBK	81.0	90.9	87.8
Random Forest	90.4	91.1	89.2
Ensemble classifier	94.2	94.7	92.7

Next, a comparative analysis is performed with the proposed methodology and existing state of art systems of depression detection using speech. Although, existing state of art methods employ different databases, features and classification, to indicate the robustness of the proposed method for depression detection the comparison is performed.

4 Comparative Analysis

A comparative analysis of the accuracies using different methods and classifiers are summarized in Fig. 4. The authors Song, Shen and Valstar [26] used SVM and CNN classifiers for the DAIC-WOZ depression database to achieve an accuracy of 85% and 86.6% respectively. Deep learned spectral features were used for the classification. Whereas Paula Lopez and Laura Docio-Fernandez [27] used a text embedding method called GloVe for depression detection. Automatic Speech Recognition on this database using GloVe method gave an accuracy of 82.9%. The comparison of the existing work with the proposed work indicates the advantage of using ensemble voting classifier. Ensemble voting classifier with and without feature selection achieved better results than the existing work.

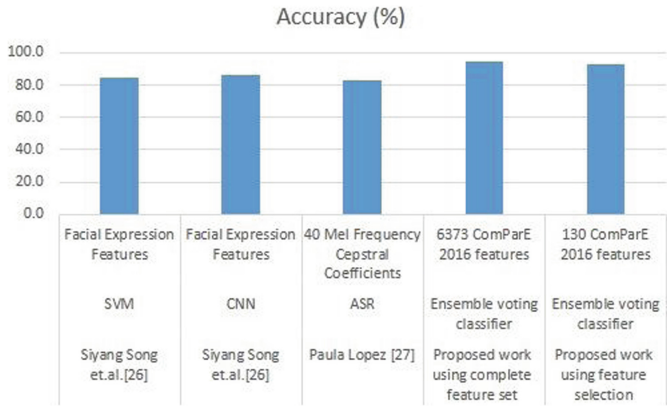

Fig. 4. Existing systems and proposed model for depression detection

5 Conclusion and Future Scope

Depression detection from speech is performed for gender dependent and independent systems wherein ComParE 2016 features and AVEC 2011 features are extracted using OpenSMILE open source toolkit. Segmentation of the speech samples of equal length on DAIC-WOZ speech database was implemented. ComParE 2016 features and AVEC 2011 features were extracted from each segment. Classification was performed using IBK, SGD, Random Forest classifiers and an ensemble voting classifier. The performance is evaluated using the accuracy obtained from each of these classifiers. The ensemble voting classifier gives better prediction results than the individual classifiers, for both the ComParE feature set as well as AVEC feature set. For speaker dependent and independent samples with and without feature selection, the performance of the ensemble classifier was superior. One of the limitation of depressive state detection is the difficulty in collecting real world speech database. In future, the method can be repeated with other database using different feature-sets. Also, the combination of ComParE 2016 and AVEC 2011 features with other relevant features can be used for depression detection. Since for speaker independent scenario, the accuracy is degraded, deep learning can be employed to obtain higher accuracy.

References

1. World Health Organization: Depression: a global public health concern (2012)
2. Scherer, S., Stratou, G., Mahmoud, M., Bober, J.: Automatic behavior descriptors for psychological disorder analysis. Image Vis. Comput. **32**, 648–658 (2014)
3. Alghowinem, S., Goecke, R., Wagner, M., Epps, J.: A comparative study of different classifiers for detecting depression from spontaneous speech. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8022–8026. IEEE (2013)

4. Moore, E.I., Clements, M., Peifer, J., Weisser, L.: Investigating the role of glottal features in classifying clinical depression. In: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 3, pp. 9–12 (2003)
5. Ooi, K.E.B., Lech, M., Allen, N.B.: Multichannel weighted speech classification system for prediction of major depression in adolescents. Biomed. Sig. Process. Control **14**(1), 228–239 (2014)
6. Ooi, K.E.B., Low, L.S.A., Lech, M., Allen, N.: Early prediction of major depression in adolescents using glottal wave characteristics and Teager Energy parameters. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, vol. 980, pp. 4613–4616 (2012)
7. Ozdas, A., Shiavi, R.G., Silverman, S.E., Silverman, M.K., Wilkes, D.M.: Investigation of vocal jitter and glottal flow spectrum as possible cues for depression and near-term suicidal risk. IEEE Trans. Biomed. Eng. **51**(9), 1530–1540 (2004)
8. Yingthawornsuk, T., Keskinpala, H.K., Wilkes, D.M., Shiavi, R.G., Salomon, R.M.: Direct acoustic feature using iterative EM algorithm and spectral energy for classifying suicidal speech. Perform. Eval. 766–769 (2007)
9. Narayanan, V., Lalitha, S., Gupta, D.: An epitomization of stress recognition from speech signal. Int. J. Eng. Technol. (UAE) **7**, 61–68 (2018)
10. Lalitha, S., Tripathi, S.: Emotion detection using perceptual based speech features. In: 2016 IEEE Annual India Conference (INDICON), Bangalore, pp. 1–5 (2016). https://doi.org/10.1109/INDICON.2016.7839028
11. Murali Krishna, P., Pradeep Reddy, R., Narayanan, V., Lalitha, S., Gupta, D.: Affective state recognition using audio cues. J. Intell. Fuzzy Syst. **3**, 2147–2154 (2019)
12. Lalitha, S., Tripathi, S., Gupta, D.: Enhanced speech emotion detection using deep neural networks. Int. J. Speech Technol. **22**(3), 497–510 (2018). https://doi.org/10.1007/s10772-018-09572-8
13. Hantke, S., Weninger, F., Kurle, R., Ringeval, F., Batliner, A.: I hear you eat and speak: automatic recognition of eating condition and food type, use-cases, and impact on ASR performance (2016)
14. McKeown, G., Valstar, M., Cowie, R., Pantic, M., Schroder, M.: The SEMAINE database: annotated multimodal records of emotionally colored conversations between a person and a limited agent. IEEE Trans. Affect. Comput. **3**(1), 5–17 (2012)
15. Gratch, J., et al.: The Distress Analysis Interview Corpus of human and computer interviews. In: Proceedings of Language Resources and Evaluation Conference (LREC) (2014)
16. www.audacityteam.org
17. Schuller, B., et al.: Affective and behavioural computing: lessons learnt from the first computational paralinguistics challenge
18. Schuller, B., Valstar, M., Eyben, F., McKeown, G., Cowie, R., Pantic, M.: AVEC 2011–the first international audio/visual emotion challenge. In: D'Mello, S., Graesser, A., Schuller, B., Martin, J.-C. (eds.) ACII 2011. LNCS, vol. 6975, pp. 415–424. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-24571-8_53
19. https://www.audeering.com/opensmile/
20. WEKA 3 - Data Mining with open source machine learning software. https://www.cs.waikato.ac.nz/ml/weka/downloading.html
21. Narayanan, V., Lalitha, S., Gupta, D.: Stress recognition using auditory features for psychotherapy in Indian context. In: ICCSP, pp. 0426–0432 (2018)

22. Bottou, L.: Large-scale machine learning with stochastic gradient descent. In: Lechevallier, Y., Saporta, G. (eds.) Proceedings of COMPSTAT 2010. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-7908-2604-3_16
23. Aha, D., Kibler, D., Albert, M.: Instance-based learning algorithms. Mach. Learn. **6**, 37–66 (1991). https://doi.org/10.1007/BF00153759
24. Pal, M.: Random forest classifier for remote sensing classification. Int. J. Remote Sens. **26**(1), 217–222 (2005)
25. https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62
26. Song, S., Shen, L., Valstar, M.: Human behaviour-based automatic depression analysis using hand-crafted statistics and deep learned spectral features. In: 2018 13th IEEE International Conference on Automatic Face and Gesture Recognition (2018)
27. Lopez Otero, P., Docio-Fernandez, L., Abad, A., García-Mateo, C.: Depression Detection Using Automatic Transcriptions of De-Identified Speech, pp. 3157–3161 (2017). https://doi.org/10.21437/Interspeech.2017-1201