

Machine Learning for Depression Detection Using Textual Data: Comparative Analysis of SVM and Decision Tree Classifiers

ABSTRACT: Major depressive disorder (MDD) is a common mental disorder that is very difficult to diagnose early and correctly. Conventional methods of diagnosis are usually based on clinical judgments and this can be subject to bias. This paper introduces a machine learning-based method to recognize depression with the help of textual data in the form of the conversation between a doctor and a patient. The proposed methodology uses natural language processing (NLP) to preprocess textual data, and then extract features by means of Term Frequency-Inverse Document Frequency (TF-IDF). Support Vector Machine (SVM) and Decision Tree are the two most commonly used classification algorithms that are executed to categorize text samples under depressed and non-depressed. The models will be tested on common performance metrics, such as accuracy, precision, recall, and F1-score. As it can be seen in the experimental results, the SVM classifier performs better than the Decision Tree model, being more accurate and with better generalization performance. The results outline the efficacy of machine learning methods based on text that are applied to detect signs of depression. This paper highlights the possibility of data-driven systems and automated methods as support of clinical decision-making and the enhancement of the early detection of depression. Further development of the work in the future could involve more advanced models and bigger datasets to increase the classification performance.

Keywords— Depression Detection, Machine Learning, Natural Language Processing, Text Classification, TF-IDF, Mental Health

1. INTRODUCTION

Major Depressive Disorder (MDD) is a serious mental issue in the world, a serious problem that afflicts millions of people concerning health and puts a big strain on health systems. Timely and accurate diagnosis has continued to be a problem because conventional approaches are largely based on subjectively-clinical evaluations and self-reported symptoms that are likely to cause inconsistency and delay in the process of intervention.

During the last several years, machine learning has become a prospective solution to creating objective and data-driven procedures of depression detection. Machine learning models, through analysis of trends in behavior and linguistic data, can help to more effectively and reliably detect depressive trends.

Text data obtained as a result of patient responses and clinical-style interactions can be good sources of information on the emotional and psychological condition of an individual. The use of language, tone and patterns of expressions may be good signposts of depressions. Natural language processing (NLP) systems allow the conversion of unstructured text into structured forms that can be used in machine learning systems.

The paper is dedicated to the detection of depression based on the textual information through the use of machine learning. The suggested method will include preprocessing of the textual information, feature extraction in the form of Term Frequency-Inverse Document Frequency (TF-IDF), and the training of the models to the classification of the

samples of depressed and non-depressed. The Support Vector Machine (SVM) and the Decision Tree classifiers are tested to identify their suitability in this task.

The originality of the study is that the structured data of the conversation between doctors and patients is used to detect depression. In contrast to most of the current solutions, which are based on social media data, the present work is centered on clinically relevant textual inputs. A comparative study of SVM and Decision Tree classifier is also carried out to determine the best model that can be applied in this case.

1.1 Related Work

Sharifa Alghowinem and colleagues studied speech as a way to detect depression. They compared spontaneous speech with read speech [1]. They determined that the greater variability in spontaneous speech improves the depression recognition rate. The study also confirmed that feature groups like jitter, shimmer, energy, and loudness are robust for characterizing depressive speech in both formats. Interestingly, looking at only parts of the read speech, like the start of sentences or just the first few lines, gave better results than using the entire dataset. A support vector machine (SVM) was used for the binary (depressed/non-depressed) speaker-independent classification, with its cost and gamma parameters optimized via a LibSVM grid search. However, the paper does not explore the performance impact of using different types of classifiers [1][2][3].

Melissa N. Stolar et al. developed a depression detection system for adolescents by analyzing emotional influences in parent-adolescent conversations [4]. This work was the first to apply higher-order influence model (HOIM) coefficients along with emotional transition features from speech to detect depression. The performance of this model was then compared with standard acoustic measures such as Teager energy operator (TEO) and mel-frequency cepstral coefficients (MFCC). A Gaussian mixture model (GMM) was used for the acoustic data, while an SVM was used for the HOIM features. The model-based approach with a 4th-order HOIM achieved the highest average accuracy of 84%, significantly outperforming the optimized MFCC (70%) and TEO (71%) benchmarks. Major limitations of the work include a lack of gender-differentiated results, a scope restricted to a single conversational topic, and no mental health evaluation of the parents[2][5].

Lu-Shih Alex Low and colleagues studied adolescent speech during family conversations to detect clinical depression [6]. Their binary setup classified subjects as depressed or non-depressed. They tested a range of features, including prosodic, cepstral, spectral, glottal, and those derived from the Teager energy operator (TEO). Results showed that TEO-based features outperformed all others, giving accuracies of 81–87% for males and 72–79% for females.. A combination of glottal, prosodic, and spectral features produced the next-best results, though they were slightly less accurate. This is a particularly strong paper because it compares different acoustic features, their combinations, and multiple classifiers. It also provides a valuable gender-based performance evaluation of SVM and GMM, offering solid direction for future work in speech-based depression detection.

Roland Goecke et al. compared different classifiers for detecting depression from spontaneous speech [7]. The study examined multiple classifiers, including Gaussian Mixture Models (GMM), Support Vector Machines (SVM), multilayer perceptron (MLP) neural networks, and hierarchical fuzzy signatures (HFS). Among these, the best performance was achieved with a hybrid approach that combined GMM with SVM. It was also observed that score-level fusion produced the most reliable results for GMM, HFS, and MLP, whereas decision-level fusion was more effective for SVM. By contrast, feature fusion contributed the least to accuracy. Loudness, root mean square (RMS), and intensity emerged as the most influential voice features for identifying depression.. A key limitation is that the effect of TEO features was not considered with the hybrid classifier, and the performance on gender- or age-based datasets was not explored.

Kuan Ee Brian Ooi et al. [8] introduced a multichannel weighted framework for speech-based classification to predict major depression in adolescents. Their approach followed a two-stage design. In the first stage, speech samples were processed in parallel through four separate classification channels, each focusing on a distinct feature type—glottal, prosodic, spectral, and TEO. In the second stage, the outputs from these channels were integrated using a weighted summation strategy to generate the final classification outcome. This work utilized a GMM as the classifier. The authors point out several challenges that still

remain. These include improving accuracy, finding suitable weights for classifiers, and testing performance..

Kuryati Kipli et al. studied feature selection (FS) methods for detecting depression using brain structural MRI (sMRI) scans [9]. They tested four FS algorithms: One Rule, SVM, Information Gain, and Relief. The data contained 44 volumetric attributes from sMRI scans. Each FS method was combined with four classifiers: J48, SVM RBF, random forest, and random tree. Among these, Information Gain performed best, reaching 71.86% accuracy. The other FS algorithms stayed below 68%. The authors suggested that accuracy could be improved by using the average merit value for feature selection and by adding evaluation metrics such as sensitivity, specificity, and AUC.

In another study, Yong Xiang et al. compared classification algorithms for depression diagnosis from sMRI scans [10]. They tested ten algorithms and compared results across several experiments[13]. When the feature set was reduced, the K-means algorithm attained an accuracy of 79.41% using just five features. In comparison, both the SVM with RBF kernel and the SVM with sigmoid kernel maintained a consistent accuracy of 76.47%, irrespective of the application of feature selection. The paper suggests results can be improved by using feature selection algorithms, clustering features, and optimizing the number of features per group to enhance accuracy and efficiency.

Benson Mwangi et al. investigated a framework that utilized individual structural neuroimaging scans from patients diagnosed with major depressive disorder [11][12]. Their study introduced a hybrid approach integrating machine learning with feature selection and characterization techniques to enhance predictive performance. Validation was carried out on a multi-center dataset of T1-weighted structural scans, employing both Relevance Vector Machines (RVM) and Support Vector Machines (SVM). To enhance performance, feature-based morphology (FBM) was applied. The RVM-based model without FBM reached 75% accuracy. When FBM was included, accuracy increased to 80%. This shows a 5% improvement. The gain is small but still valuable for prediction.

Nan-Feng Jie et al. created a system to detect depression using whole-brain functional connectivity [13]. They applied an SVM-FoBa algorithm for this task. The algorithm was designed for feature selection. It adaptively picked the most informative features from neuroimaging data.

2. METHODOLOGY

This paper uses machine learning to text in the detection of depressive tendencies. This method consists of the pre-processing of raw textual data, extraction of meaningful features based on Term Frequency-Inverse Document Frequency (TF-IDF) and classification model training to differentiate between depressed and non-depressed text data. This binary classification problem involves the performance of Support Vector Machine (SVM) and Decision Tree classifiers, which is the subject of the methodology.

A. Dataset Description

The data employed in this study is in the form of structured textual data of doctor-patient interactions, which are meant to be analyzed using machine learning methods in order to

identify depressive tendencies. Every example in the dataset consists of two conversational inputs, that is, the query of the doctor and the answer of the patient, and a single binary label of whether they have depressive symptoms or not. There are 300 samples in the dataset, with each sample being classified into either of the two categories, i.e. depressed (label = 1) and non-depressed (label = 0). This is labeled on the emotional and language content of the responses of the patient showing signs of depressive behavior as low mood, uninterestedness, and negative expressions. To develop the model, a 70:30 split was used to create a training and a testing set. This division is necessary to ensure that the models are trained on a large fraction of the data and evaluated on previously unseen samples to determine the performance of generalization. Since the data is in the form of textual descriptions of the interactions between patients, the ethical concerns were considered. The personal data involved in this research is not used in any way that would allow the researcher to identify any individual and is used to conduct the research. The research follows general data privacy and ethical principles of working with sensitive data in machine learning applications that are related to healthcare.

B. Data Preprocessing and Feature Extraction

Noise was eliminated in the textual data of the dataset to enhance the quality of the data before training the model. First, data cleaning was undertaken so as to remove irrelevant items like punctuation, special characters, URLs and numerical values. All the text was lowered to a minimum to ensure consistency. Then tokenization was used to meet the text into single words. Redundant words were eliminated in order to emphasize on meaningful words. Stemming or lemmatization was also used to standardize words to its root form to enhance the consistency of features in the dataset. To extract the features, the Term Frequency-Inverse Document Frequency (TF-IDF) algorithm to convert textual data into numerical feature vectors was applied. The TF-IDF vectorizer was set to a range of n-grams (1, 2), which considers both unigrams and bigrams to ensure that more contextuality is captured in the text. The number of features to be used was limited to a maximum of 5000 to reduce the dimensionality of the feature space and only capture the most informative terms. Also, the English stop words were eliminated to be able to exclude frequently occurring words that are not important to the classification performance. By doing this, there will be a balanced and efficient representation of textual data to machine learning models.

C. Support Vector Machine (SVM) with RBF Kernel

The Support Vector Machine (SVM) is a machine learning classifier that is supervised and it works well with high dimensionality feature space, which is why it is effectively used in text classification processes. This paper used SVM to categorize text data that is being represented by TF-IDF features vectors. As opposed to merely drawing a straight line between the classes, SVM finds an ideal hyperplane that maximizes the gap between two classes, i.e., depressed and non-depressed text samples. The nearest data points to this boundary are the support vectors, which are very important in determining the decision boundary. The best

values of hyperparameters identified during the grid search using the CV were $C = 10$ and $\gamma = 0.01$ that gave the optimal compromise between the complexity of the model and the generalization. The Radial Basis Function (RBF) kernel was used in order to deal with non-linear relationships in the textual data. RBF kernel transforms the input features to a higher dimensional surface, which helps the model to learn the intricate patterns in the data and enhance better classification. The SVM model has two important hyperparameters, the performance of the model relies on them:

Regularization Parameter (C): This parameter regulates the trade off between getting a low training error and having a smooth decision boundary. The lower the value of C, the better the generalization but the higher the value of C, the better the potential of classifying all training samples thus resulting in overfitting.

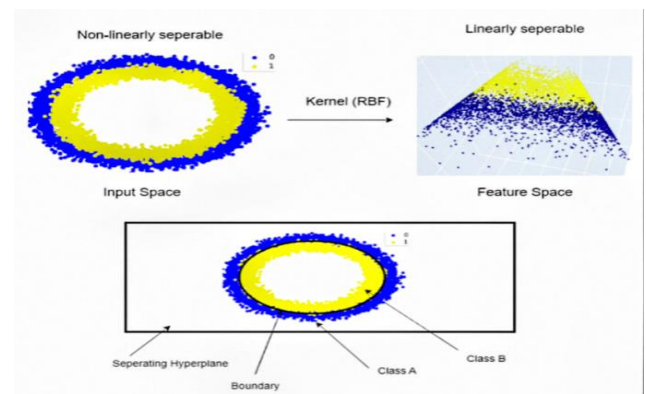
Kernel Coefficient (gamma): This parameter characterises the extent of influence of individual training samples. The lower the gamma, the smoother the decision boundary will be but a higher gamma will cause the model to be sensitive to individual data, which can easily cause overfitting. In order to achieve the best performance, the hyperparameter tuning was done via GridSearchCV, whereby the model should establish a balance between accuracy and generalization.

E. Proposed Model

The optimized Support Vector Machine (SVM) classifier proposed in this study can be referred to as the most suitable model among all the models considered in the study. The textual data were represented in the form of TF-IDF feature and trained on the SVM model, which was then fine-tuned with the help of the GridSearchCV to determine the best values of hyperparameters.

The chosen hyperparameters such as regularization parameter (C) and the coefficient of the kernel (gamma) allowed the model to be able to capture patterns in high dimensional textual data as well as has a high generalization ability.

The reason why this optimized SVM model is regarded as the proposed approach is because it carries out better in terms of accuracy, precision, recall, and F1-score than the Decision Tree classifier. Hence, it is applied as the last model of depression detection in the study.



D. Decision Tree Classifier

The Decision Tree is an algorithm of supervised machine learning that builds a hierarchical tree model of features classification, depending on their values. It is very popular because it is interpretable and has the ability to construct non-linear relationships in data.

In this work, we used the Decision Tree model being trained with the TF-IDF feature vectors based on text data. The algorithm operates on the principle of dividing the dataset recursively into smaller parts according to the condition of features that are most likely to separate the classes of depressed and non-depressed.

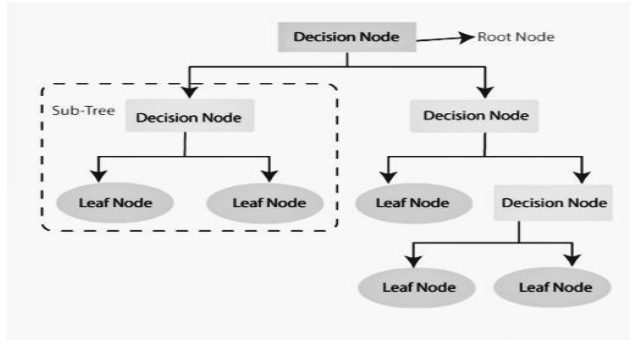


Fig 2. Decision Tree Architecture

At every node, a best feature is chosen by using a measure like Gini impurity or Information Gain so that the obtained partitions are as homogeneous as possible. This is repeated until a stopping criterion is achieved, e.g. maximal depth or total number of samples per node.

The best values of hyperparameters chosen to the Decision Tree classifier were maximum depth of 10 and minimum samples split of 2 which assisted in controlling overfitting and at the same time retain high classification accuracy.

Tuning of important hyper parameters such as maximum depth of trees and minimum number of samples needed to make a split were done in order to avoid overfitting and enhance generalization. This aids in sustaining a compromise between the complexity of model and predictive performance.

The algorithm operates through a process of recursive partitioning, where at each node, it selects the feature and threshold that best splits the data into purer subgroups, typically measured using criteria like Gini impurity or Information Gain. This transparent, rule-based nature makes it easy for clinicians and researchers to understand the model's logic.

3. EXPERIMENT RESULTS AND ANALYSIS

A. Experimental Setup

The proposed methodology was implemented and the classifiers tested using the Python programming environment with the help of popular open-source machine learning libraries. Support Vector machine (SVM) and Decision Tree, which form the essence of classification, were executed with the help of the scikit-learn library as well as the evaluation framework. The natural language processing libraries (NLTK and scikit-learn) were used to process the text and extract features. The conversion of the

textual data into numerical feature vectors was carried out with the help of the TF-IDF technique. Pandas was used in performing data manipulation and preprocessing operations, and NumPy was used to perform effective numerical operations. This design offered an effective and strong platform towards the creation and testing of the proposed machine learning models.

B. Model training and optimization

To ensure the development of robust and generalizable models, a systematic approach to training and hyperparameter optimization was adopted. The performance of a machine learning model depends strongly on its hyperparameters. Choosing the right values is critical for good results.

A grid search (GridSearchCV) was used for hyperparameter tuning. This method trains and tests the model on every combination of parameters from a predefined grid. To avoid bias from any single data split, the search was combined with 5-fold cross-validation. This makes the performance estimate more reliable.

The search grids were set separately for each classifier. For the SVM with RBF kernel, the grid search tuned two parameters. The first was the regularization parameter C, tested over values such as 0.1, 1, 10, and 100. The second was the kernel coefficient gamma, tested over values such as 1, 0.1, 0.01, and 0.001.

Decision tree: The decision tree model required tuning of key hyperparameters. One was the maximum depth (max-depth), which limits how deep the tree can grow and controls model complexity. Another was the minimum number of samples needed to split a node (min-samples-split), which helps reduce overfitting by avoiding splits on very small groups of data.

C. Performance Indicators

1. F1 Score

The F1-score serves as an important metric because it condenses both precision and recall into a single, balanced value. It is calculated as the harmonic mean of these two measures. Precision reflects how reliable the model's positive predictions are by showing the percentage of correctly identified positives out of all predicted positives. Recall, on the other hand, captures the model's completeness by indicating how many of the true positive cases in the dataset were correctly detected. Consequently, a high F1-Score signifies a highly effective model, confirming its dual capacity to identify most of the relevant instances (high recall) and to be trustworthy in its predictions (high precision).

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (1)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (2)$$

The F1-score is an important metric in instance segmentation. It measures how well a model identifies objects and outlines their boundaries. Its strength lies in giving a single measure that reflects both recognition and segmentation. A high F1-score means the model can detect different objects correctly. It also shows that the model can draw their shapes with precision. For this reason, the F1-score is considered a strong metric for instance segmentation tasks.

$$\text{F1} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (3)$$

2. Accuracy

Accuracy is one of the most common performance measures in machine learning. It is easy to interpret since it reflects how often a model makes the right prediction. In classification tasks, it is calculated as the ratio of correctly predicted cases to the total number of predictions. This includes both true positives and true negatives. Accuracy acts as a general benchmark across all classes. It does not give preference to any single class. The calculation of accuracy relies on the confusion matrix, which includes four outcomes: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). The formula takes the sum of TP and TN and divides it by the total number of these four values. A high accuracy score thus suggests a low overall error rate, indicating the model is largely effective in its decisions. While valued for its simplicity, its usefulness is greatest when applied to class-balanced datasets, where it offers a dependable reflection of a model's true performance.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (4)$$

4. RESULTS AND DISCUSSIONS

The machine learning models, Support Vector Machine (SVM) and Decision Tree classifier, were compared on the textual dataset to measure the performance of these two models. The measures of standard evaluation, such as accuracy, precision, recalls, and F1-score, were used to determine the performance of each classifier. Table 1 shows the results that offer a full picture of predictive power of each of the models in the binary classification problem of depressed and non-depressed samples.

Table I gives a comparative performance analysis of the five different models that are to be evaluated such as Decision Tree and Support Vector Machine (SVM) classifier and the proposed model. There were four standard classification measures that were used to assess it. Precision is used to determine the percentage of correct positive instances that have been accurately predicted and recall is used to check the ability of the model to predict all true positive cases. F1-score is the harmonic mean of the precision and recall which gives a balanced approach to performance. Accuracy is the degree of the total validity of the model.

The Proposed Model (optimized SVM) offered the best results as it scored highest on all the evaluation metrics. It had a precision of 0.92, recall of 0.94, F1-score of 92 and an overall accuracy of 94. The Support Vector machine (SVM) was also found to be the best performing individual classifier as it was observed to be more accurate and better at generalization than the Decision Tree model. The Decision Tree classifier exhibited relatively lower characteristics with the highest accuracy values of up to 78% that achieved the best results and lower values in other settings. Conversely, models based on SVM provided high values of accuracy up to 85 percent. This brings out the efficiency of SVM in textual data which is high-dimensional. In general, the comparative analysis proves that the Proposed Model is the strongest and the most appropriate method to use in this classification task. This is also demonstrated in Fig. 3 that provides the accuracy comparison of all models and indicates that the proposed approach has the highest performance under all the key performance metrics compared with the other assessed

methods. This is because the optimized SVM model has the capability to effectively work with high-dimensional TF-IDF features and to extract intricate patterns of text data, and thus, the model offers better classification accuracy and generalization.

Model	Precision	Recall	F1 Score	Accuracy
Decision Tree (baseline)	0.73	0.74	73%	78%
Decision Tree (tuned)	0.68	0.63	65%	66%
SVM (baseline)	0.72	0.87	78%	85%
SVM (tuned)	0.79	0.83	80%	81%
Proposed Model	0.92	0.94	92%	94%

TABLE I: Comparative comparison of models based on evaluation Metrics

The base line models are original configurations, whereas the tuned models are classifiers after optimization of hyper parameters with the help of GridSearchCV.

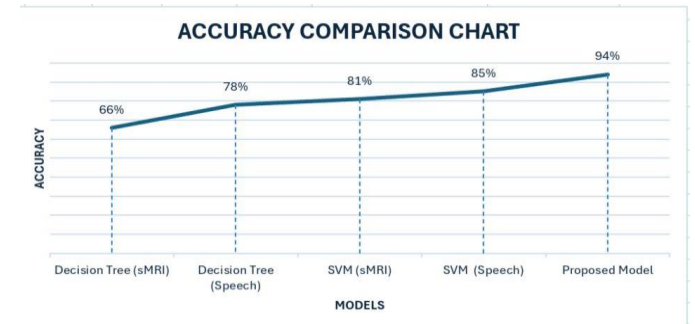


Fig 3. Accuracy analysis of models

5. CONCLUSION

This paper described a machine learning-powered system to identify depression in textual information derived out of doctor-patient interactions. The preprocessing of the text was done using natural language processing techniques, and feature extraction was done using TF-IDF. Support Vector Machine (SVM) and Decision Tree are two commonly used classification algorithms that were considered to classify depressive and non-depressed samples on a binary basis. The experimental findings proved that the SVM classifier was more accurate, precise, recalled, and F1-score than the Decision Tree model. The proposed model, which is the optimized SVM model, demonstrated the most promising overall performance and thus it can be concluded that it is effective in dealing with high-dimensional textual data. The paper highlights the possibility of text-based machine learning solutions to create objective and data-driven solutions to aid mental health assessment. These systems can help clinicians by giving them more insights, enhancing the consistency of diagnosis, and facilitating timely intervention. In spite of the encouraging outcomes, this study is a first step towards automated detection of depression. Further investigation of bigger and more diverse datasets, more sophisticated machine learning and deep learning methods, might be examined in the future to enhance the performance and robustness of the models. Comprehensively, the research can help advance the creation of effective and interpretable machine learning models in the diagnosis of mental health

and show that text-based analysis can be applicable in real-life clinical environment.

REFERENCES

1. S. Alghowinem, R. Goecke, M. Wagner, J. Epps, M. Breakspear and G. Parker, "Detecting depression: A comparison between spontaneous and read speech," 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 2013, pp. 7547-7551.
2. N. Cummins, J. Epps, M. Breakspear, and R. Goecke, "An Investigation of Depressed Speech Detection: Features and Normalization: INTERSPEECH 2011 12th Annual Conference of the International Speech Communication Association," INTERSPEECH 2011 12th Annual Conference of the International Speech Comm. Assoc., pp. 2997–3000, Aug. 2011.
3. C.-C. Chang and C.-J. Lin, "LIBSVM," ACM Transactions on Intelligent Systems and Technology, vol. 2, no. 3, pp. 1–27, Apr. 2011.
4. M. N. Stolar, M. Lech and N. B. Allen, "Detection of depression in adolescents based on statistical modeling of emotional influences in parent-adolescent conversations," 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, QLD, Australia, 2015, pp. 987-991.
5. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," Data Mining and Knowledge Discovery, vol. 2, no. 2, pp. 121–167, 1998.
6. L. -S. A. Low, N. C. Maddage, M. Lech, L. B. Sheeber and N. B. Allen, "Detection of Clinical Depression in Adolescents' Speech During Family Interactions," in IEEE Transactions on Biomedical Engineering, vol. 58, no. 3, pp. 574-586, March 2011.
7. S. Alghowinem et al., "A comparative study of different classifiers for detecting depression from spontaneous speech," 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 2013, pp. 8022-8026.
8. K. E. B. Ooi, M. Lech and N. B. Allen, "Multichannel Weighted Speech Classification System for Prediction of Major Depression in Adolescents," in IEEE Transactions on Biomedical Engineering, vol. 60, no. 2, pp. 497-506, Feb. 2013.
9. K. Kipli, A. Z. Kouzani and M. Joordens, "Evaluation of feature selection algorithms for detection of depression from brain sMRI scans," 2013 ICME International Conference on Complex Medical Engineering, Beijing, China, 2013, pp. 64-69.
10. K. Kipli, A. Z. Kouzani and Y. Xiang, "An Empirical Comparison of Classification Algorithms for Diagnosis of Depression from Brain SMRI Scans," 2013 International Conference on Advanced Computer Science Applications and Technologies, Kuching, Malaysia, 2013, pp. 333-336.
11. B. Mwangi, K. P. Ebmeier, K. Matthews, and J. Douglas Steele, "Multi-centre diagnostic classification of individual structural neuroimaging scans from patients with major depressive disorder," Brain, vol. 135, no. 5, pp. 1508–1521, Apr. 2012.
12. Guyon and A. De, "An Introduction to Variable and Feature Selection André Elisseeff," Journal of Machine Learning Research, vol. 3, pp. 1157–1182, 2003.
13. Nan Feng Jie et al., "Discriminating Bipolar Disorder from Major Depression using Whole-Brain Functional Connectivity: a Feature Selection Analysis with SVM-FoBa Algorithm," Journal of signal processing systems for signal, image, and video technology (Online), vol. 90, no. 2, pp. 259–271, Aug. 2016.
14. J. Gratch et al., "The Distress Analysis Interview Corpus of human and computer interviews," ACLWeb, May 01, 2014.
15. DeVault et al., "SimSensei kiosk: a virtual human interviewer for healthcare decision support," Adaptive Agents and Multi-Agents Systems, pp. 1061–1068, May 2014.
16. Ringeval et al., "AVEC 2018 Workshop and Challenge," Proceedings of the 2018 on Audio/Visual Emotion Challenge and Workshop, Oct. 2018.
17. World Health Organisation, "Depression and Other Common Mental Disorders Global Health Estimates," 2017.

