Detecting Depression from Human Conversations

Jesia Quader Yuki
North South University
Dhaka, Bangladesh
jesia.quader@northsouth.edu

Md. Mahfil Quader Sakib
North South University
Dhaka, Bangladesh
mahfil.sakib@northsouth.edu

Zaisha Zamal
North South University
Dhaka, Bangladesh
zaisha.zamal@northsouth.edu

Sabiha Haque Efel
North South University
Dhaka, Bangladesh
sabiha.haque@northsouth.edu

Mohammad Ashrafuzzaman Khan
North South University
Dhaka, Bangladesh
mohammad.khan02@northsouth.edu

ABSTRACT
Depression is the silent killer of the new age. Although depression is considered a mental illness it can affect physical health. Most often people ignore it until physical or acute mental symptoms start showing up. One way to early detect symptoms of depression is to analyze what people are talking about - their conversations. This is an attractive solution as it is much less noninvasive than conventional medical tests and has the potential to be quite accurate. In this paper, we explored this idea. We used transcripts of conversation data and analyzed the text using ten different machine learning algorithms including LSTM, SVM, Ensemble methods, Random forest, and Decision tree. As our goal is to find any patterns in spoken language people used to express depression - screams, cheers, mumbles, whines, stutters, murmurs, etc., we applied the learning algorithms on original transcripts of the conversations. We found that LSTM performed best (about 94% accuracy) in finding depressed dialogues followed by naïve Bayes (about 69.05% accuracy).

CCS Concepts
• Applied computing → Health care information systems • Computing methodologies → Information extraction.

Keywords
Depression; human conversations; sequence learning; LSTM; NLP.

1. INTRODUCTION
Depression has become the most common and prevalent mental health issue around the world. Advancements and adoption of technologies like the internet, smartphones, social networks, etc., into human life is making the situation even worse [1-4].

Although primarily considered a mental issue, depression can cause heart and coronary diseases, sleep issues, troubles with memory & decision making, social withdrawals & isolations, weakened immune systems, etc. However, with proper treatments on time, these disasters could be easily avoided. However, most people do not take depression as a serious reason to visit doctors.

If not treated properly and on-time, people become sad and try to get comfort in drugs, alcohol addictions, etc. It changes people from what they (actually) are, their views of society and other people. These sometimes become the main cause of unthinkable social disasters like suicides, mass shooting incidents [22], etc.

There are many ways to detect and address depressions. In the last couple of years, researchers addressed this problem from many interesting directions. Some most common ways are biomarkers, brain imaging & signals, facial & speech analysis, social network content analysis, etc.

However, we took a different approach – we used the text form of human conversations. When people use formal conversations, like writing social network status, blogs, etc., they either keep some social etiquette or sometimes exaggerate too much which removes important emotions and clues from the conversations. However, for everyday conversations with friends, colleagues, neighbors, bosses, etc., most often there is a balance. That is the reason we became interested to explore this interesting direction.

In this research, we used the dataset [23]. This dataset contains data for both depressed and non-depressed participants. We converted the conversation to text form. Then we used different kinds of machine learning and deep learning techniques to find if the speaker is depressed or not. We used LSTM, KNN, Naive Bayes, SVM, Logistic Regression, Decision Tree, Random Forest, AdaBoost, Bagging, and Extra Tree methods for the classification. We found that LSTM performed better than any other method on a large scale. The other methods performed almost the same.

The rest of the papers is organized as follows: Section 2 describes the related works, Section 3 describes the dataset we used in this research, Section 4 describes the steps and methodology used, Section 5 describes the results, Section 6 discusses future directions of this research, and finally, Section 7 concludes this paper.

2. RELATED WORKS
Researchers have approached this problem from many different directions. Most of the research used statistics and machine learning at some point. The main difference is the data source used to detect depression.

ICCCM'20, July 17–19, 2020, Singapore, Singapore

ACM ISBN 978-1-4503-8766-8/20/07…$15.00
https://doi.org/10.1145/3411174.3411187

2.1 Using Online Social Network Content

Many researchers used the content posted on social networks for analyzing depression. Despande et al [5] applied Support Vector Machine (SVM) and Multinomial Naive Bayes (MNB) on data collected from social media like Twitter, Facebook, and Instagram. They found that MNB could classify with 83% accuracy which SVM did with an accuracy of 79%. Islam [7] et al analyzed social network data using k-nearest neighbor, decision tree, ensemble methods, and support vector machines. Instead of finding the accuracy, they studied the precision and recall values. They found that among these three methods support vector machines performed worse. Their [7] main goal was to analyze social media content and find out the statistics on how much people share depressing content and about their distribution over the timeline. Nguyen et al [18] used naive Bayes, support vector machines, logistic regression for classifying online posts of depression communities. However, they found that linguistic features are important in classifying posts into depression and other classes.

Therefore, most works in this direction used content shared by the social network users and tried a bunch of machine learning and statistical methods to correctly classify the content as depressed or not. All these works found evidence that the text data shared by the users contain information about depression. However, the data shared in social networks are either too diplomatic or too exaggerated. Therefore, in this work, we explore another direction – the transcription of the voice data.

2.2 Using Brain Imaging & Signals

Some research used brain imaging, signals, etc., to classify depression. Hosseinifard [6] et al used the EEG signal to find out if the patient is depressed or not. Then they extracted features from the four bands using detrended fluctuation analysis, higuchi fractal, correlation dimension, and lyapunov exponent. They used k-nearest neighbor, linear discriminant analysis, and logistic regression for the classification. They achieved at most 90% accuracy using the logistic regression method. The other two methods could achieve at most 83.3% accuracy. Acharya et al [13] also used the EEG signal to find the presence of depression. They also used a support vector machine, k-nearest neighbor, naive Bayes, neural networks, and decision tree methods. They achieved 90% accuracy using a support vector machine classifier. Patel et al [19,20] used imaging and different statistical processing to detect depression.

Bestehera et al [11] tried to find a correlation between brain age and depression. However, they found no significant correlation between brain-age and depression. McGinnis et al [12] studied depression and anxiety on children by analyzing voice data. They mainly analyzed the audio features of the voice data while we used the natural language processing techniques on the text extracted from audio data. And, unlike them, our study is concentrated on adults, not children.

Kipli et al [9] analyzed structural MRI brain data for detecting depression. They extracted several features from the structural MRI data: volume calculated from the whole brain, white matter, grey matter, and hippocampus. They found that a combination of support vector machines, expectation-maximization, and random tree-based methods could reach at almost 85.23%.

However, collecting these imaging and signal data requires visiting doctors, going to hospitals, access to special hardware, and expensive for most patients. Therefore, our approach is cheaper and more accessible to common people.

2.3 Using Facial and Voice Analysis & Biomarkers

Some researchers analyzed facial and voice analysis, biomarkers, etc., to classify and find depression. Cohn et al [14] used facial expression and vocal prosody to detect depression. They used a support vector machine and logistic regression for the classification. they achieved 88% accuracy from the facial expression and 79% accuracy from the vocal prosody.

Dipnall et al [8] used logistic regression to find biomarkers that could identify the presence of depression. The final three biomarkers they found very effective were red cell distribution width, serum glucose, and total bilirubin. They found significant interactions between total bilirubin with Mexican American/Hispanic groups and current smokers. They further [16] worked on connecting factors especially environmental factors to depression using unsupervised self-organized maps and supervised boosted regression methods.

These methods are either less accurate or still require visiting doctors, hospitals, testing, etc. Our approach is non-invasive and more suitable for common people.

2.4 Other Approaches

Lee et al [17] studied the use of different machine learning techniques in Ovid Medline/PubMed publications and their accuracies. They found that combinations of methods always performed better than single methods.

Hamouda et al [24] proposed an SVM based method to find significant words for sentiment analysis. Their main contribution is to use SVM to classify the sentiment words from the amazon reviews. Guo [25] proposed a framework to analyze students' sentiment from their social media posts. They used the content of ratemyprofessor.com for analyzing the students' sentiment about professors. However, for the sentiment analysis, the author used online sentiment analysis services provided by IBM. Both [24] and [25] used written text for their work while we used the transcribed voice communication data.

Iliou et al [10] developed a preprocessing method for depression type prediction for mentally ill patients. The point of their work was to classify the patients according to their type of depression. For that, they used several preprocessing techniques: principal com technique, ILIOU preprocessing strategy, and evolutionary search technique. Then they used several classification algorithms: the nearest neighbor classifier, random forest, multi-layer perceptron, support vector machine, and fuzzy logic.

The goal of this research is to find out the suitability of using machine learning for depression detection. However, instead of just showing general approaches, we used the transcription of voice data to classify depressed speakers.

3. DATASET DESCRIPTION

We used an interview dataset [23] for this research. The dataset contains interviews of various kinds of people – male/female, different ethnicity, depressed/non-depressed, anxious/calm, patients with PTSD (post-traumatic stress disorder)/non-PTSD.

The interviews were conducted by humans, human-controlled agents, and autonomous agents. In this way, the authors of [23] collected audio, video, and questionnaire responses. Then they transcribed and annotated parts of the corpus for a variety of verbal and non-verbal features.

Therefore, the corpus contains data from the most important human communications means. Face-to-face interviews attempted to simulate talking to friends, family members, strangers, and colleagues. The teleconference simulated talking over the phone to known/unknown people. The fully automated interviews simulated asking questions to smart devices.

When asked the same question, the method of the interview had notable effects on the answers. For example, if the people were asked "who has the most positive influence on life" face-to-face, they mostly picked their spouses. Interestingly, they usually selected parents if asked by computers. Again, people talked about helpfulness, loyalty, trust as the positive sides in their spouses while talked about being loving/caring as the positives in their parents. The corpus seems to have a realistic distribution of daily human conversations done in different situations.

For our research, we used 15604 episodes of conversations from this corpus. We used the text from the transcribed dialogues. The used dataset contains 63% normal (non-depressed) and 37% depressed dialogues.

4. METHODOLOGY

In this paper, we took the standard machine learning approach to classify depression. However, when we implemented, we used all the standard software-engineering best practices so that we can use any algorithms while testing now and in the future. The stages shown in figure 1, are all divided into separate modules.

The first step was to find the important words. We kept the data almost original. Just cleaned up some for unusual punctuation marks. We kept the words having the potential to express emotion.

At first, humans tagged a dataset as depressed or not from the conversation. Each example consists of one complete sentence spoken by a human.

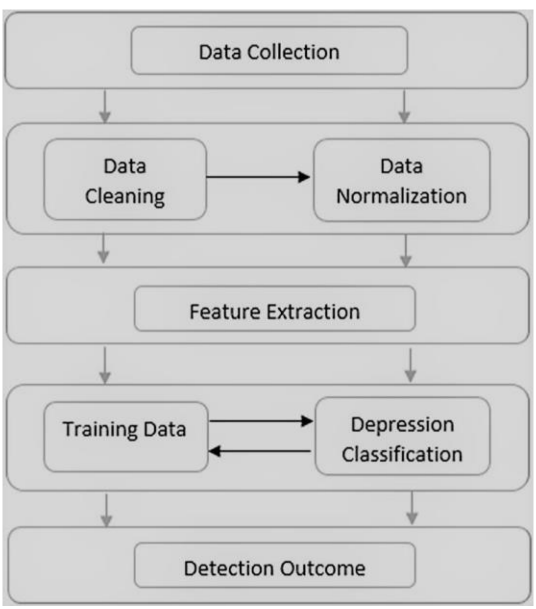

Figure 1. Classification process.

In this paper, we evaluated LSTM, KNN, Naive Bayes, SVM, Logistic Regression, Decision Tree, Random Forest, and different ensemble methods such as AdaBoost, Bagging, and Extra Tree algorithms.

5. RESULTS

Table 1 shows the results from all the algorithms applied in this paper.

However, before using all the algorithms, we wanted to analyze the text. Our goal was to find indicating words for classifying depression. For that, we used a bag-of-words model and made a list of words most used by depressed people and most used by non-depressed people. Overall, most frequent words were "love", "little", "um", "probably" etc. Surprisingly, depressed people use the word "life", "feel", etc., more than the normal people. Depressed people are also more likely to be concerned and worried about life, thus use negative words in the same sentences (containing the word "life"). Interestingly, no conversation by normal (non-depressed) people contains the word "feel". Non-depressed people used the word "love" frequently. However, for depressed people, the word "love" was not in the most frequent 40 words spoken! Both classes are observed to use the word "um" at the highest rate and we found no significant correlation. Therefore, we got rid of "um" during the cleaning process.

Table 1. Algorithms used and their accuracy.

Algorithm	**Accuracy**
LSTM	93.95 %
SVM	63.63%
KNN	56.34%
Naive Bayes	69.05%
Logistic regression	66.60%
Decision tree	63.39%
Random Forest	67.02%
AdaBoost	64.57%
Bagging	65.76%
Extra Tree	66.45%

Then we used the standard machine learning techniques on the bag of words. The naïve Bayes performed best and classified 69.05% accurately. After that, we used a support vector machine. We found that SVM could achieve at most 63.63% accuracy. After that, we tried several other algorithms as shown in figure 2. All of them achieved almost the same level of accuracy.

However, we were not happy with these results as we found evidence by manual check that there is enough information in the data for the classification.

Then we used a technique, LSTM [21], that is capable of exploiting the relative position of the words – a sequence learning technique. Long short-term memory (LSTM) uses neural networks with feedback connections and takes the position of the words into account. Normally, LSTM is widely used for natural language processing.

We found dramatic improvements in accuracy after using LSTM, we achieved 93.95% accuracy. This proves our intuition, the latent presence of subtle emotion carried into the sequence of words was found using the right tool – the LSTM network.

LSTM takes the lead here. However, Naïve Bayes and Random Forest have similar accuracy nonetheless much less than the LSTM. To compare the Naïve Bayes and Random Forest, we took the precision and recall values for these two methods. The precision and recall values for Naïve Bayes are 0.63 and 0.37; for Random Forest they are 0.88 and 0.39. Therefore, we see that Random Forest has much higher precision than the Naïve Bayes.

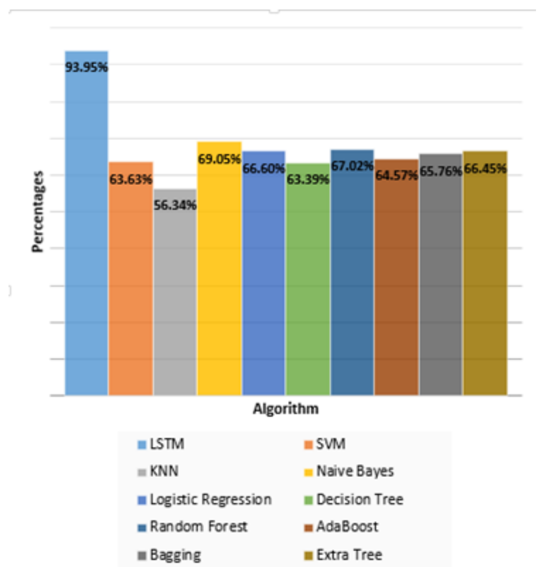

Figure 2. Accuracy comparison for ten different methods.

6. FUTURE WORKS

We want to further explore this problem using mundane day-to-day, completely informal conversation. The main problem with that is to tag the data. However, we are working to build such a dataset using non-intrusive smartphone apps.

The app could contain the depression detection model and warn the users of imminent dangers if the symptoms are ignored for a long time. It will also be very interesting to explore the possibility, advantages, and disadvantages of integrating such apps with real medical data visible by the primary care doctors.

7. CONCLUSION

What we found in this research is that machine learning algorithms are powerful in detecting depression in human conversations. Especially, one of the recurrent neural networks models – LSTM achieved almost 94% accuracy. It proves our intuition that the words used during a conversation have important information to classify the speech as depressed or not. However, other machine learning methods failed to achieve such a high level of accuracy as they do not take the sequence into account.

8. REFERENCES

[1] Cotten, S.R., Ford, G., Ford, S. and Hale, T.M., 2012. Internet use and depression among older adults. *Computers in human behavior*. 28, 2 (Mar. 2012), 496-499. DOI= 10.1016/j.chb.2011.10.021

[2] Yen, J.Y., Yen, C.F., Wu, H.Y., Huang, C.J. and Ko, C.H., 2011. Hostility in the real world and online: the effect of internet addiction, depression, and online activity. *Cyberpsychology, Behavior, and Social Networking*. 14, 11 (Apr. 2011), 649-655. DOI= 10.1089/cyber.2010.0393

[3] Thomée, S., Härenstam, A. and Hagberg, M., 2011. Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults-a prospective cohort study. *BMC public health*. 11, 1 (Feb. 2011), 66. DOI= 10.1186/1471-2458-11-66

[4] Błachnio, A., Przepiórka, A. and Pantic, I., 2015. Internet use, Facebook intrusion, and depression: Results of a cross-sectional study. *European Psychiatry*. 30, 6 (May. 2015), 681-684. DOI= 10.1016/j.eurpsy.2015.04.002

[5] Deshpande, M. and Rao, V., 2017. Depression detection using emotion artificial intelligence. *In 2017 International Conference on Intelligent Sustainable Systems (ICISS). IEEE*, (Dec. 2017), 858-862. DOI= 10.1109/ISS1.2017.8389299

[6] Hosseinifard, B., Moradi, M.H. and Rostami, R., 2013. Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal. *Computer methods and programs in biomedicine*. 109, 3 (Oct. 2013), 339-345. DOI= 10.1016/j.cmpb.2012.10.008

[7] Islam, M.R., Kabir, M.A., Ahmed, A., Kamal, A.R.M., Wang, H. and Ulhaq, A., 2018. Depression detection from social network data using machine learning techniques. *Health information science and systems*. 6, 1 (Dec. 2018), 1-8. DOI= 10.1007/s13755-018-0046-0

[8] Dipnall, J.F., Pasco, J.A., Berk, M., Williams, L.J., Dodd, S., Jacka, F.N. and Meyer, D., 2016. Fusing data mining, machine learning and traditional statistics to detect biomarkers associated with depression. *PloS one*. 11, 2 (Feb. 2016). DOI= 10.1371/journal.pone.0148195

[9] Kipli, K., Kouzani, A.Z. and Hamid, I.R.A., 2013. Investigating machine learning techniques for detection of depression using structural MRI volumetric features. *International journal of bioscience, biochemistry and bioinformatics*. 3, 5 (2013), 444-448. DOI= 10.7763/IJBBB.2013.V3.252

[10] Iliou, T., Konstantopoulou, G., Ntekouli, M., Lymperopoulou, C., Assimakopoulos, K., Galiatsatos, D. and Anastassopoulos, G., 2019. ILIOU machine learning preprocessing method for depression type prediction. *Evolving Systems*. 10, 1 (Oct. 2019), 29-39. DOI= 10.1007/s12530-017-9205-9

[11] Besteher, B., Gaser, C. and Nenadić, I., 2019. Machine-learning based brain age estimation in major depression showing no evidence of accelerated aging. *Psychiatry Research: Neuroimaging*. 290 (Jun. 2019), 1-4. DOI= 10.1016/j.pscychresns.2019.06.001

[12] McGinnis, E.W., Anderau, S.P., Hruschak, J., Gurchiek, R.D., Lopez-Duran, N.L., Fitzgerald, K., Rosenblum, K.L., Muzik, M. and McGinnis, R.S., 2019. Giving Voice to Vulnerable Children: Machine Learning Analysis of Speech Detects Anxiety and Depression in Early Childhood. *IEEE journal of biomedical and health informatics*. 23, 6 (Apr. 2019), 2294-2301. DOI= 10.1109/JBHI.2019.2913590

[13] Acharya, U.R., Sudarshan, V.K., Adeli, H., Santhosh, J., Koh, J.E., Puthankatti, S.D. and Adeli, A., 2015. A novel depression diagnosis index using nonlinear features in EEG signals. *European neurology*. 74, 1-2 (Aug. 2015), 79-83. DOI= 10.1159/000438457

[14] Cohn, J.F., Kruez, T.S., Matthews, I., Yang, Y., Nguyen, M.H., Padilla, M.T., Zhou, F. and De la Torre, F., 2009. Detecting depression from facial actions and vocal prosody. *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*. IEEE. (Sep. 2009), 1-7. DOI= 10.1109/ACII.2009.5349358

[15] Dipnall, J.F., Pasco, J.A., Berk, M., Williams, L.J., Dodd, S., Jacka, F.N. and Meyer, D., 2016. Fusing data mining, machine learning and traditional statistics to detect biomarkers associated with depression. *PloS one*. *11*, 2 (Feb. 2016). DOI= 10.1371/journal.pone.0148195

[16] Dipnall, J.F., Pasco, J.A., Berk, M., Williams, L.J., Dodd, S., Jacka, F.N. and Meyer, D., 2017. Why so GLUMM? Detecting depression clusters through graphing lifestyle-environs using machine-learning methods (GLUMM). *European Psychiatry*. 39 (Jan. 2017), 40-50. DOI= 10.1016/j.eurpsy.2016.06.003

[17] Lee, Y., Ragguett, R.M., Mansur, R.B., Boutilier, J.J., Rosenblat, J.D., Trevizol, A., Brietzke, E., Lin, K., Pan, Z., Subramaniapillai, M. and Chan, T.C., 2018. Applications of machine learning algorithms to predict therapeutic outcomes in depression: A meta-analysis and systematic review. *Journal of affective disorders*. 241 (Aug. 2018), 519-532. DOI= 10.1016/j.jad.2018.08.073

[18] Nguyen, T., O'Dea, B., Larsen, M., Phung, D., Venkatesh, S. and Christensen, H., 2017. Using linguistic and topic analysis to classify sub-groups of online depression communities. *Multimedia tools and applications*. 76, 8 (Apr. 2017), 10653-10676. DOI= 10.1007/s11042-015-3128-x

[19] Patel, M.J., Andreescu, C., Price, J.C., Edelman, K.L., Reynolds III, C.F. and Aizenstein, H.J., 2015. Machine learning approaches for integrating clinical and imaging features in late‐life depression classification and response prediction. *International journal of geriatric psychiatry*. 30, 10 (Feb. 2015), 1056-1067. DOI= 10.1002/gps.4262

[20] Patel, M.J., Khalaf, A. and Aizenstein, H.J., 2016. Studying depression using imaging and machine learning methods. *NeuroImage: Clinical*. 10 (Nov. 2016), 115-123. DOI= 10.1016/j.nicl.2015.11.003

[21] Gers, F.A., Schmidhuber, J. and Cummins, F., 2000. Learning to forget: Continual prediction with LSTM. 12, *Technical Report IDSIA-01-99.* 10 (Oct. 2000), 2451-71. DOI= 10.1162/089976600300015015

[22] Metzl, J.M. and MacLeish, K.T., 2014. Mental illness, mass shootings, and the politics of American firearms. *American journal of public health*. 105, 2 (Dec. 2014), 240-249. DOI= 10.2105/AJPH.2014.302242

[23] Gratch, J., Artstein, R., Lucas, G.M., Stratou, G., Scherer, S., Nazarian, A., Wood, R., Boberg, J., DeVault, D., Marsella, S. and Traum, D.R., 2014. The distress analysis interview corpus of human and computer interviews. *In Language Resources Evaluation Conference.* (Jan. 2014), 3123-3128.

[24] Hamouda, A., Marei, M., & Rohaim, M., 2011. Building Machine Learning Based Senti-word Lexicon for Sentiment Analysis. *Journal of Advances in Information Technology*. 2, 4 (Nov. 2011), 199-203. DOI= 10.4304/jait.2.4.199-203

[25] Guo, D., 2015. Tracking Student Sentiment from Social Media. *Journal of Advances in Information Technolog*y. 6, 2 (Jan. 2015), 80-83. DOI= 10.12720/jait.6.2.80-83