Towards Detecting ADHD and other Comorbid Conditions from Social Media Posts

Akhil Pillai
ap6863@nyu.edu

Abstract

Information present in social media are often used in performing sentiment analysis. This technique has also helped in identifying potential mental health problems from user posts or responses on this platform. However, like many other mental health problems, Attention Deficit Hyperactivity Disorder (ADHD) is a complex disorder that have symptoms that overlap other disorders. Hence, detecting the presence of the condition along with other comorbid conditions associated with ADHD can be a very challenging task. To address this, we propose using multi-label classifiers to identify ADHD and related conditions based on the information available online. In this project, we will train two multi-label classification models, one based on LSTM and the other on BERT, and analyze and compare their performance metrics with existing approaches on the same.

1 Introduction

Social media has always been an outlet to many users for sharing their thoughts, emotions and feelings. Information present on these sites have often been used to perform sentiment analysis by analyzing the language, tone, and context used in user posts, comments or replies (Go et al. (2009)). This work has even been extended to perform analysis on text to look for the presence of mental health problems and have shown promising results in using binary classifiers ((Coppersmith et al., 2014), (De Choudhury et al., 2013)) to detect the presence of such problems. However, one of the major challenges of using such classifiers is the issue of comorbidity[1]. Conditions like ADHD often have many symptoms overlapping with one or the other conditions such as Anxiety, Depression, Bipolar Disorder, Obsessive Compulsive Disorder etc. While using binary classifiers may be a good starting point in identifying these conditions, the approach is not suitable to narrow down the more complex and deeper issues present in users.

Multi-Label classifiers can prove to be more suitable to such tasks, where it can assign multiple labels simultaneously to indicate the presence of one or multiple disorders. In this paper, we attempt to identify ADHD and its potential comorbid conditions Anxiety, Depression and OCD using multi-label classifiers. To perform multi-label classification, we propose two multi-label models, one based on Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997) and the other on Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019). These models will be trained on the user posts scraped from subreddits such as r/ADHD, r/Anxiety, r/Depression and r/OCD. The results of these experiments show that for smaller datasets the proposed LSTM and BERT models outperform the previous binary classification models. Our study also highlights the importance of having good quality labelled data for training such models. The key takeaway from our study is that using multi-label classifiers can potentially be the next step in identifying mental health problems (besides ADHD) and understanding the complex relationships with other mental health conditions.

The rest of the paper is structured as follows. In Section 2, we highlight previous work done in this area and usage of binary classifiers. In Section 3, we provide a detailed description of our proposed models. In Section 4, we describe our data collection process. In Section 5 we present our experimental setup, results and analysis. Finally, in Section 6, we discuss our findings and suggest directions for future research.

2 Related Work

Social media has been a popular platform for users to express their thoughts and feelings, and re-

[1] Comorbidity refers to the presence of two or more medical or psychiatric conditions in a patient simultaneously.

searchers have explored the use of this data for sentiment analysis and mental health identification. In particular, binary classifiers have been commonly used to detect the presence of mental health issues such as depression and anxiety based on social media data.

For instance, in a study by (Coppersmith et al., 2014), Twitter data was used to train a binary classifier for identifying users with depression. The classifier was able to achieve an accuracy of 70% in detecting depression. Similarly, in a study by (De Choudhury et al., 2013), Facebook data was used to develop a binary classifier for identifying users with depressive symptoms, with an accuracy of 86%.

However, the use of binary classifiers has limitations in identifying more complex and overlapping conditions such as comorbidities. Some studies have explored the use of multi-label classifiers to solve this challenge. For example, in a study by (Ren et al., 2018), a multi-label classification model was proposed to identify mental health conditions using social media data. The model was able to identify multiple mental health conditions simultaneously, achieving an average F1-score of 0.51.

Overall, while previous work has explored the use of social media data for mental health identification, the limitations of binary classifiers in detecting comorbidities have led to a need for more such multi-label classifiers.

3 Approach

We propose two multi-label classifiers: one based on LSTM the other on BERT. The model design and details are explained further in the subsections that follow.

3.1 LSTM

Figure 1 shows the architecture of the LSTM model. The text input is tokenized using pre-trained GloVe embeddings (Pennington et al., 2014) and then fed into the embedding layer. Then it goes to the LSTM layer which is a single layer with 64 nodes. The output of the LSTM layer is sent into an output dense layer which has the same number of nodes as the output labels, 4. The output layer uses the sigmoid activation function to produce probabilities for each label.

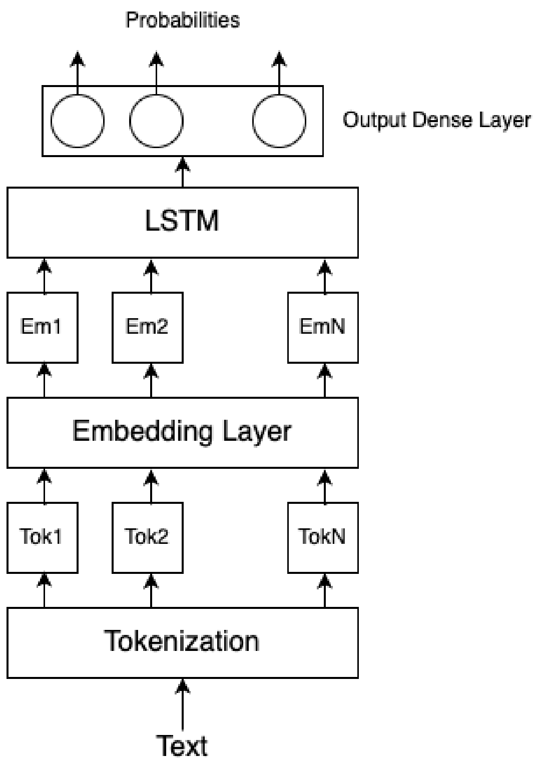

Figure 1: Architecture of our proposed LSTM model.

3.2 BERT

Figure 2 shows the architecture of the BERT model. The text input is tokenized using the pre-trained BERT tokenizers. The tokenized input is then sent to the pre-trained BERT model (bert-base-uncased) from the Hugging Face Transformers library. Just like the LSTM model, it has an output dense layer with 4 nodes, each representing a label and a sigmoid activation function to produce probabilities.

4 Dataset

In this section, we describe the process we used to collect data for training our models. This process is necessary since there is limited data on people diagnosed with ADHD; and even then it is not accessible publicly since such data is usually medical records and diagnosis of patients.

4.1 Collection

We followed a similar approach to (Kim et al., 2020), scraping data from subreddits. Reddit has several subreddits[2] dedicated to different topics of interest including mental health conditions. Of these subreddits we choose four of them, namely

[2]Subreddit is a specific community within reddit where users can post and discuss content related to a particular topic or interest

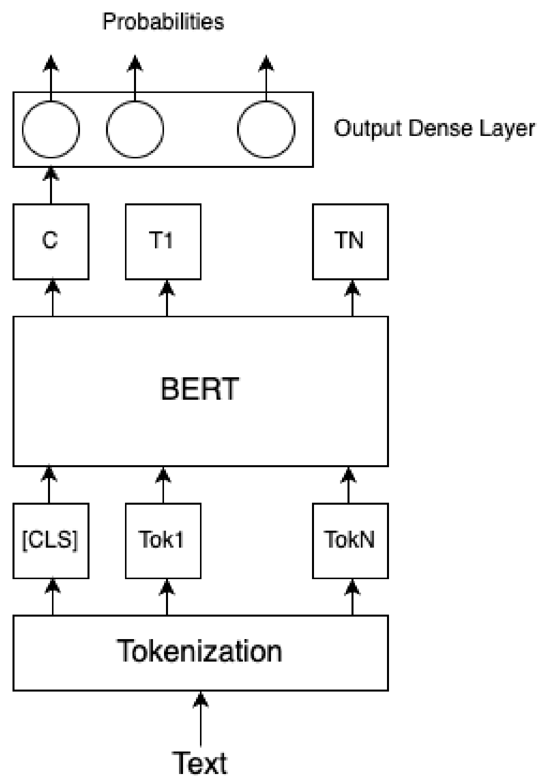

Figure 2: Architecture of our proposed BERT model.

r/ADHD, r/Anxiety, r/depression and r/OCD. Each of these subreddits have posts by users, which usually describes a difficulty a user is facing due to the said condition, or venting about having this conditions in general or just simply seeking support from others. We collected the top posts of each subreddit over a period of two weeks.

4.2 Cleaning

In total we collected over 10k posts of users. To protect the identity of the users and to respect their privacy we discarded any personally identifiable data and only considered the post itself. We further filtered posts that were flagged 'deleted' or has a word length shorter than 80 characters or longer than 3000 characters.

4.3 Labelling

Each post has four features, with each feature describing the presence of the mental health condition. For example, if the post belongs to the subreddit r/ADHD, the ADHD feature will be labelled as 1 and the same process is repeated over other subreddits. We use information from flairs[3] and crossposts[4] to label other features. For example, if a post belongs to r/ADHD and has been crossposted to r/depression, then we can label this as both ADHD and depression.

[3]Flairs are used to categorize or label posts.

Subreddit	Posts	Word Count/Post
r/ADHD	2986	312
r/Anxiety	2758	396
r/depression	2890	355
r/OCD	2778	226
Overall	11412	322

Table 1: Subreddits and the number of posts scraped for each.

4.4 Description

The dataset is a csv file with a collection of posts and their features. Each feature is either a 0 or 1 representing the absence or presence of each of the four conditions present in the post by the user.

Table 1 shows the statistics of the data collection process.

5 Experiments

In this section we present the experiments and the results for the classification task on both the models.

5.1 Setup

We use the dataset described in the previous section and create copies of its subsets with varying sizes (100, 200, 300, .. 1000, .. posts). For the LSTM model, we trained the model for a maximum of 100 epochs, with a learning rate of 0.001 and batch size of 32. For the BERT model, we fine-tuned the pre-trained model for a maximum of 20 epochs with a learning rate of 0.00001 and dropout of 0.1. We used F-Score as our metric to evaluate the performance of our models.

5.2 Results

Figures 3 and 4 show the performance of our proposed models on varying dataset sizes. As seen in the charts, the models performed differently for different sizes of training and testing data. For smaller dataset sizes (< 200 posts), the models performed very well reaching high F-score values. For relatively larger dataset sizes (200 - 800 posts), the performance reached its peak but then dropped

[4]Crosspost is a post that has been shared to a subreddit from another subreddit.

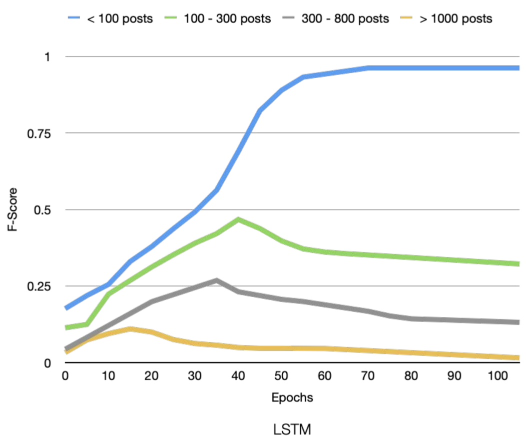

Figure 3: Performance of the proposed LSTM model.

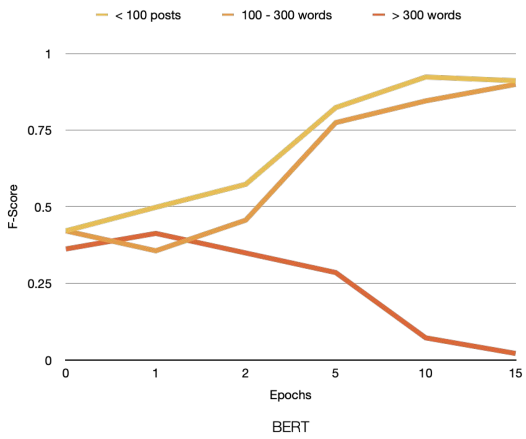

Figure 4: Performance of the proposed BERT model.

on further training. For dataset sizes even larger (> 800 posts), the performance dropped extremely reaching nearly zero F-scores on training.

Table 2 shows the maximum obtained F-scores for each model for the respective dataset sizes.

5.3 Analysis

There are several explanations for the high performance of both models on smaller datasets. One reason is that the smaller datasets were relatively more uniform in terms of their content and structure, which could have reduced the prevalence of outliers or noise, ultimately leading to greater learning and generalization capabilities for the models. Furthermore, a smaller dataset size can facilitate faster and more efficient training, thus allowing the models to converge more quickly and potentially yield better performance outcomes. As the size of the dataset increased, the performance of the models dropped, as the models encountered more and more noisy data. Since the data is scraped from reddit and labeled programatically, the quality of the data and the associated labels is not the best either.

Another factor to be considered for the drop in performance is the use of the embeddings and tokenization methods (GloVe, BERT) used, which may not be suitable or fine-tuned for the text data available on reddit. The number of dimensions in the embedding and LSTM layers could attribute to not capturing essential features (or overfitting). However, without the availability larger higher quality data, it is difficult to narrow down on a single reason for drop in performance of these models.

No. of Posts	LSTM	BERT
< 100	0.962	1.000
100 - 200	0.483	0.955
300 - 500	0.282	0.366
500 - 800	0.197	0.086
> 800	0.089	0.018

Table 2: Highest F-score values obtained for the models trained on different dataset sizes.

6 Conclusion

In this paper, we proposed and presented multi-label classification models in an attempt to detect ADHD and the potential comorbid conditions that come with it. While our results did not demonstrate an improvement to the binary classifiers trained on each condition, it did present promising results when trained on limited data. With the availability of better quality labeled data on mental health conditions, we believe that our approach holds promise for future research in this area. Overall, we believe that our work provides an important foundation for future research in using sentiment analysis to detect mental disorders not limited to ADHD and we look forward to further exploring this promising avenue for improving mental health diagnosis and treatment.

Acknowledgements

We would like to express our gratitude to the moderators on reddit for providing forums for multitude of users that post seeking support to their mental health problems, and providing a means to extract the posts using the API. We also would like thank our professor He He for providing the opportunity to dive into this topic and attempt to find a solution

to mental problems universal to all.

References

Glen Coppersmith, Mark Dredze, and Craig Harman. 2014. Discovering fine-grained sentiment with the STanford TWITTER SENTIMENT CORPUS. In *Proceedings of the Eighth International Conference on Weblogs and Social Media*, pages 11–20. AAAI Press.

Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In *Seventh International AAAI Conference on Weblogs and Social Media*. Citeseer.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.

Alec Go, Richa Bhayani, and Lei Huang. 2009. Sentiment analysis on twitter data. *Journal of Computational Science*, 2(1):1–8.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation*, 9(8):1735–1780.

Jina Kim, Jieon Lee Lee, Eunil Park Park, and Jinyoung Han. 2020. A deep learning model for detecting mental illness from user content on social media. *Journal of Information Science*, 46(4):459–473.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

Yanqing Ren, Han Jiang, Lei Wang, and Yijun Liu. 2018. Multi-label classification of mental health conditions via social media data. In *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 505–506. IEEE.