Automatic Detection and Classification of Cognitive Distortions in Journaling Text

Mai Mostafa, Alia El Bolock and Slim Abdennadher
German University in Cairo, Egypt

Keywords: Cognitive Distortions, Cognitive Behavioral Therapy, Mental Health, Machine Learning, Deep Learning, Natural Language Processing.

Abstract: Cognitive distortions are negative thinking patterns that people adopt. Left undetected, it could lead to developing mental health problems. The goal of cognitive behavioral therapy is to correct and change cognitive distortions that in turn help with the recovery from mental illnesses such as depression and anxiety, overcoming addictions, and facing common life challenges. The aim of this study is to provide a machine learning solution for the automatic detection and classification of common cognitive distortions from journaling texts. Relatively few works have focused on exploring machine learning solutions and tools in the context of cognitive-behavioral therapy. And, given the rising popularity of online therapy programs, this tool could be used for instant feedback, and would also be a helpful service for therapists and psychiatrists to initiate and ease the detection of cognitive distortions. In this study, we provide a novel dataset that we used to train machine learning and deep learning algorithms. We then employed the best-performing model in an easy-to-use user interface.

1 INTRODUCTION

Cognitive distortions describe the dysfunctional core beliefs and misconceptions a person might have, that control the way people feel towards themselves and the world around them. These maladaptive cognitions highly influence the way people react emotionally, psychologically, and how they behave (Beck, 2011). For example, “The plant I just got died, I will never have a beautiful garden because everything will die” is a type of cognitive distortion, because it reached a conclusion about a single isolated negative event, and applied that conclusion on all future plants. Cognitive distortions are commonly grouped into 15 types (Beck, 1976). However, there is no evidence-based way to classify cognitive distortions. And it’s important to recognize that there is a degree of overlap between them. Moreover, a single sentence can exhibit multiple types of cognitive distortions. For example, “I failed this interview, I’ll probably fail all interviews I get” can be classified as overgeneralization, as well as magnification, and catastrophizing. For these reasons, we have decided to pick only a couple of types of cognitive distortions for the purpose of this study. Definitions and examples of the cognitive distortions covered in this study are provided in table 1 (de Oliveira, 2012).

In many cases, cognitive distortions result in feelings such as anxiety and depression. Beck’s cognitive theory for depression suggests that people with inaccurate and negative core beliefs are more susceptible to depression. This cognitive theory is based on the grounds that an individual’s affect and behavior are largely determined by the way in which they structure the world (Beck, 1987). Cognitive Behavioral Therapy (CBT) is a therapeutic approach that is derived from Cognitive Therapy model theory (Beck, 1976; Beck, 1987) that helps patients recognize and identify their own thinking errors and distorted view of reality. They are then helped to correct these thinking errors, and are taught cognitive and behavioral skills so that they can develop more accurate beliefs and adopt a healthier way of making sense of the world around them. CBT was attributed to help with the treatment of anxiety disorders, somatoform disorders, bulimia, anger control problems, and general stress (Hofmann et al., 2012). This approach holds people accountable to their own thoughts and feelings, and rather than only delve into the past to know the reasons for their thought fallacies, the goal is to

Mostafa, M., El Bolock, A. and Abdennadher, S.
Automatic Detection and Classification of Cognitive Distortions in Journaling Text.
DOI: 10.5220/0010713000003058
In *Proceedings of the 17th International Conference on Web Information Systems and Technologies (WEBIST 2021)*, pages 444-452
ISBN: 978-989-758-536-4; ISSN: 2184-3252

Table 1: Cognitive distortions, definitions and examples.

	Cognitive distortions	Definitions	Examples
1	Overgeneralization	I take isolated cases and generalize them widely by means of words such as “always”, “never”, “everyone”, etc.	“Every time I have a day off from work, it rains.” “You only pay attention to me when you want sex”.
2	Should statements (also “musts”, “oughts”, “have tos”)	I tell myself that events, people’s behaviors, and my own attitudes “should” be the way I expected them to be and not as they really are.	“I should have been a better mother”. “He should have married Ann instead of Mary”. “I shouldn’t have made so many mistakes.”

identify and correct them. Recently, online therapy programs have gained a lot of popularity. These programs are developed to accompany, or replace in-person CBT (Ruwaard et al., 2012) One of the main reasons that make it unique and important is because it can be more frequently accessed, which was found to be one major component for the effectiveness of CBT and leads to a more rapid recovery (Bruijniks et al., 2015).

This study is conducted to develop methods for the automatic detection and classification of cognitive distortions found in mental health journals. It will be of assistance to therapists in online therapy programs. Providing detection and instant feedback and allowing them to scale more easily. Only a few machine learning studies were conducted in relation to mental health. Fewer in the context of cognitive behavioral therapy. The goal of this study is to collect a novel dataset to be used to explore ways to detect and classify cognitive distortions, and provide machine learning and deep learning methods for the detection and classification of two common cognitive distortions. As well as develop a user interface to visualize the performance of the tool and put it to use. Which would be highly beneficial and easy for therapists to use in online therapy programs.

2 RELATED WORK

2.1 Data Collection

There is a wide variety of choices when it comes to data collection. Most papers studying sentiment analysis and emotion recognition have used already existing datasets that are publicly available to conduct their research. Unfortunately, due to the fact that cognitive distortion detection and classification is still not widely researched, we haven’t been able to find an available dataset. In this subsection, we discuss multiple sources for data collection.

2.1.1 Crowdsourcing

Crowdsourcing platforms are used as a means to collect data from a large group of paid participants. For the purpose of collecting texts portraying cognitive distortions, participants are given a brief description of a cognitive distortion, then asked to mention a situation or event, where they have exhibited that type of thinking (Shickel et al., 2019).

2.1.2 Online Therapy Partnerships

datasets have been collected in partnership with Koko (Morris et al., 2015). Which is an online therapy program that is based on peer-to-peer therapy. As well as TAO, an online therapy program implemented in various universities across the USA. As part of the program, students are requested to fill out journals and logs to track their progress. Texts collected from actual journals are argued to be a more accurate representation of the cognitive distortion than those collected by crowdsourcing. Since the authors of those text passages weren’t specifically asked to recall a situation where they exhibited a certain way of thinking (Shickel et al., 2016; Shickel et al., 2019; RojasBarahona et al., 2018).

2.1.3 Social Media APIs

Social media and Twitter in particular is an ideal platform to collect data from. As it provides texts with the same natural expression of cognitive distortions as those in journals. Meaning that the authors of the texts are not asked to specifically recall a situation where they felt they were thinking in a specific manner. In addition to the easy, free of charge use of the application programming interface (API), it can provide big volumes of data in a short amount of time. Due to the popularity of the platform itself, and the ease of data collection, many academic research studies have employed the

Twitter API to build their dataset. (Hu et al., 2019) (Mozeticˇ et al., 2016)(Cliche, 2017)(Chatterjee et al., 2019). (Campan et al., 2018) Have shown that using Twitter API is a reliable way of collecting data for research purposes.

2.2 Methods for Detection and Classification

Cognitive distortion detection and classification tasks are similar to the tasks of emotion detection and sentiment analysis. In a way, emotion classification and cognitive distortion classification are tasks to classify different negative sentiments. We have compiled and referred to a few studies in these areas in this section.

2.2.1 Rule-based Approach

Rule-based knowledge consists of grammatical and logical rules to follow. The approach may rely on dictionaries, lexicons, and ontologies.

Keyword Recognition: The task is to find occurrences of certain keywords in a sentence. These keywords are stored in a constructed dictionary or lexicon.(Bracewell et al., 2006) presented an emotion dictionary, where emotion words and phrases were gathered from different sources including news articles. These words were then labeled either positive or negative. An emotion classification algorithm is then used on news articles to classify the overall sentiment. The algorithm counts the number of positive and negative emotion words, and a simple equation is used to determine the article's emotion.

Ontological Knowledge: Gruber defined an ontology as "an explicit specification of a conceptualization"(Gruber, 1993). Ontologies offer meaning to terms and address the relationship between them. Most medical ontology applications follow a symptom-treatment or symptom-diagnosis categorization. Some are used to assist health professionals in clinical decisions by making evidence-based inferences. These inferences are delivered by providing knowledge through the ontology regarding treatments, symptoms, diagnosis, and prevention methods(Yamada et al., 2020), therefore require limited options for input. Nonetheless, ontologies were used to assist with natural language processing (NLP) applications when it comes to categorizing a natural language text, or with Artificial Intelligence (AI) chatbots. One such ontology is introduced in (Estival et al., 2004) as part of a virtual environment project. Where the NLP unit receives input from the user and builds a natural language query. The reasoning subsystem with the help of the ontology evaluates the query and delivers a natural language answer. (Shiv-hare and Khethawat, 2012; Minu and R.Ezhilarasi, 2012) were able to classify emotions from natural language texts based on an emotion hierarchy defined by the ontology. Ontologies are also utilized to understand and recognize the way of speaking when feeling a certain emotion, and to get the similarity between sentences, not just to classify the emotion based on keywords (Haggag et al., 2015).

2.2.2 Learning-based Approach

Traditional Learning: The automatic detection and classification of emotions from texts are in great demand. A lot of papers have studied multiple approaches and techniques to be able to perform such a task. One of the methods is classifiers such as Support Vector Machine (SVM) that are trained to be able to detect emotions (Teng et al., 2006)(Balabantaray et al., 2012)(Hasan et al., 2014). (Asghar et al., 2020) applied and compared different machine learning algorithms, which are Na¨ıve Bayes, Random Forest, Support Vector Machine (SVM), Logistic regression, K-Nearest neighbor, and XG boost to try and suggest the algorithm with the best text classification results. The algorithm that performed best with respect to the accuracy, recall, and precision was the logistic regression algorithm. Detecting and classifying cognitive distortions is an important task for the improvement of online therapy services. Both tasks of detecting whether a text contained cognitive distortions or not, and classifying a text known to contain a cognitive distortion into one of fifteen cognitive distortions have been performed. After testing out multiple classifiers, it was found that logistic regression performs best for a relatively small data set (Shickel et al., 2019).

Deep Learning: Given a large data set, deep learn-ing techniques can outperform and scale more effectively with data, than traditional machine learning techniques. In addition, given the fact that it requires less feature extraction and engineering, it is increasingly being adopted for natural language processing tasks. One such task is SemEval 2017 task 4. Which includes Twitter sentiment classification on a 5-point scale (Rosenthal et al., 2017). The best performing system belonged to (Cliche, 2017) which uses Long Short-Term Memory (LSTM) and Convolutional Neural

Network (CNN) models. For the participation of (Baziotis et al., 2018) in SemEval 2018 Task 1, which included determining the existence of none, one or more out of 11 emotions in Twitter texts. Bidirectional LSTM were trained by a fairly large data set of around 60,000 annotated tweets. LSTM models were also used by (Cachola et al., 2018) who focused on the effect of using vulgar words and expressions on the perceived sentiment.

Using a large data set, deep learning models were trained, and unsupervised learning for a large quantity of unlabeled data was utilized to classify cognitive distortions, as well as emotions and situations (RojasBarahona et al., 2018).

3 METHODS

3.1 Data Collection and Annotation

Due to the fact that cognitive distortion detection and classification tasks are not widely researched topics, there is no publicly available dataset containing text with labeled cognitive distortions. Hence, we collected and annotated a novel dataset. The dataset contains text passages labeled into one of three categories. Namely, overgeneralization, should statement, and non-distorted. A summarization of the dataset is provided in table 2. Each collected entry was reviewed for relevance and annotated by the authors and a life coach with a Meta coaching certification. The life coach was presented with the text data in a shared excel sheet. The sheet contained the sentences, the given label, and a checkbox. There was another column next to the checkbox that was left blank to be filled with the correct label in case the given label was incorrect. Corrections to the dataset were applied according to the excel sheet.

Twitter API: We decided to collect data from Twitter. The social media platform provides an easy-touse API that can be deployed to collect big volumes of data in a short amount of time. Using the API, we only collected the body of the tweet, no demographics or any other information about the author of the tweet were collected. Search words were required for filtering relevant tweets. From the examples provided by (de Oliveira, 2012), we have been able to deduce a pattern or form that sentences exhibiting a certain cognitive distortion usually acquire. One example, "Every time I have a day off from work, it rains" the sentence form that could be derived is "Every time . . . , it . . . " Where something negative happens after "it". Overall, 1122 entries were collected using the API, and they were reviewed for relevance and labeled.

Web Crawling: Examples of cognitive distortions are provided on most websites and blogs about cognitive behavioral therapy. we collected some of these examples, as well as examples provided in research papers. (Beck, 1970; Yurica and DiTomasso, 2005; de Oliveira, 2012).

Survey: We also constructed and distributed a survey. We first presented the participants with a short description of the cognitive distortion and provided two examples. We then asked the participants to recall a time in their own lives when they exhibited the described pattern of thinking, and provide examples of what they might have said to themselves, or to others. We encouraged participants to provide multiple examples or paraphrase the same example. The survey was distributed on different social media platforms, and participants were requested to share it. In total, we were able to collect 147 entries from 49 responses. These responses were reviewed for relevance and labeled.

HappyDB Dataset: We utilized (Asai et al., 2018) data set to collect non distorted texts. HappyDB was collected using crowdsourcing, where the workers were asked to answer either: "what made you happy in the last 24 hours?" or, "what made you happy in the last 3 months?" We added 1101 answers to our dataset and labeled them as nondistorted. These entries were again reviewed for relevance. It's important for the research to collect nondistorted texts, as the goal is to create a tool that can automatically detect cognitive distortions. So providing plenty of nondistorted examples was crucial to be able to separate distorted and nondistorted texts.

Preprocessing: We performed common preprocessing techniques, including converting all text to lower case and removing punctuation and emojis. For the machine learning models, a couple of vectorizers were used. Namely, tf-idf vectorizer, and count vectorizer. These vectorizers transformed our dataset textual entries into sparse vectors. Multiple n-gram ranges were tested using these vectorizers, to find that, in general, unigrams and bigrams performed the best. We also utilized multiple dense embeddings that are most popular in similar NLP tasks for the machine learning models, such as GloVe, Bert, and Flair. For our deep learning models, we train 100 and 300 dimensions for GloVe embeddings, as well as BERT embeddings.

Table 2: Summary statistics for the dataset.

	Non-distorted	Over-generalization	Should statements
Twitter API	178	518	426
Web crawling	—	18	21
Survey	—	65	82
HappyDB	1101	—	—
Total	1279	601	529

3.2 Models

We define our task to be the ability for a model to distinguish between nondistorted text, and text containing one of two cognitive distortions. This task creates an all-inclusive model for the detection and classification of two common cognitive distortions. This is important from a mental health point of view because it can alert the practitioner to the presence of cognitive distortions, and guide the patient's treatment options. We experimented with multiple machine learning models. Including logistic regression (LR), support vector machines (SVM), and Naive Bayes (NB). As mentioned in the preprocessing part of section 3.1, features were extracted via term frequency-inverse document frequency (tf-idf) vectorizer, or count vectorizer. We also experimented with different word embeddings for the LR and SVM models. Optimal hyperparameters were tuned via grid search and included model regularization and solvers. Convolutional neural networks (CNN), and long short-term memory (LSTM) were applied to construct the deep learning models. The architectures of the CNN and LSTM models can be seen in figures 1 and 2 respectively. We perform an 80/20 split of the data to train and test sets, setting the random state to a constant to ensure the same train and test sets for every model. Three layers of CNN along with their max pooling were applied. We used filter windows of 3, 4, and 5. These layers were then concatenated and flattened. A dropout layer was added, then a dense layer. For the LSTM model, a spatial dropout layer is placed after the embedding layer and before the LSTM layer with a drop rate of 0.2. For both models, we tuned hyperparameters by trying different values for each hyperparameter. We set the best performing value of one hyperparameter before tuning the next one. The results of these experiments are discussed in section 4.

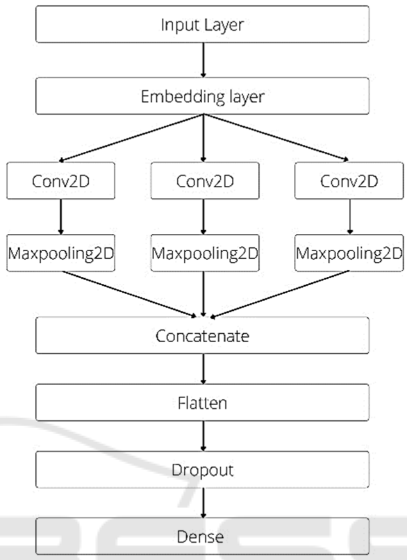

Figure 1: Proposed CNN model architecture.

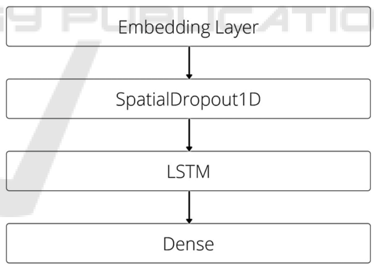

Figure 2: Proposed LSTM model architecture.

3.3 User Interface

Functionality: The purpose of developing a user interface (UI) is to create human interaction with the model. This UI is intended for psychiatrists, therapists, or life coaches. Who receive journals from patients via online therapy programs, or in any other electronic way. Once the model is provided with text passages, it goes through the passage sentence by sentence, automatically detecting and classifying cognitive distortions in the text. If any cognitive distortion is detected, the sentence that represents one of the cognitive distortions in the text would be highlighted with a certain color. The user is informed what color belongs to what distortion. This makes the user instantly aware of the presence of cognitive distortions in the text. This tool is very easy to use, saves time when it comes to detecting cognitive distortions, and ensures that no cognitive distortions will be left undetected. The website first presents the user with instructions on how to use the tool, as well as a color map for highlighting the cognitive distortions. A text box is provided for the user to enter text passages. Once the text is submitted, it gets copied on the side of the text box, with the sentences that contain cognitive distortions highlighted in color. No information submitted through the website is saved in any way.

Development: We developed this website using the Django framework. Django framework is a python-based free and open-source web framework. Figure 3 demonstrates the architecture of the website. The input text provided by the user is preprocessed and vectorized using the same techniques as the data in the dataset when the model was being trained. The model can be easily loaded onto the folder where the website is being developed. This gives plenty of room for model improvements and updates. Once the model is loaded onto the script, it can be used for classification, provided preprocessed and vectorized text. The input text is displayed for the user highlighted with the color associated with the cognitive distortion, or not highlighted at all in case the sentence didn't exhibit any cognitive distortions. The project after development was deployed on Heroku, a container-based cloud Platform as a Service (PaaS), with the domain www.cognitivedistortion-detection.herokuapp.com.

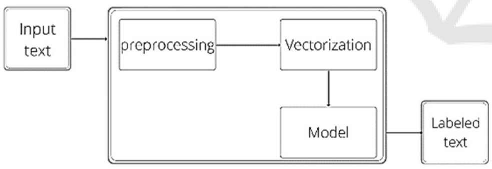

Figure 3: Proposed UI architecture and flow.

4 RESULTS

In this section, we report the performance of both our machine learning and deep learning models. Our task consists of the detection and classification of two types of cognitive distortions. In our dataset, each entry is labeled as nondistorted, contains overgeneralization, or should statement. Our dataset has a noticeably higher number of nondistorted examples than examples containing a cognitive distortion due to the assumption that texts containing a certain cognitive distortion mostly share a number of keywords or sentence structures. Unlike nondistorted verbal expressions that have wider ranges of sentence structures and expressions. As mentioned in section 3.2, we experimented with different vectorizers to extract features for different n-gram ranges, to find that for both vectorizers, unigrams and bigrams resulted in the best performance. We attribute this to the common use of words or small sequences of words in texts containing cognitive distortions. An example would be the common use of the words "Never will" or "Always" when overgeneralization is being expressed. In table 3, we report the precision, recall, and F1 scores for the machine learning models. It's clear to see that logistic regression and SVM models perform almost the same. Both yield an F1 score of 0.95. We attribute this to the similar nature of the algorithms. We also include comparable results obtained from training with different word embeddings. BERT embeddings performed the best, yielding an F1 score of 0.93, with weighted precision and recall of 0.93 for both.

Table 3: Machine learning models results.

Model	Precision	Recall	F1
LR-count-vectorizer	0.95	0.95	0.95
LR-BERT	0.93	0.93	0.93
LR-Flair	0.87	0.87	0.87
LR-GloVe	0.82	0.82	0.82
SVM-count-vectorizer	0.95	0.95	0.95
SVM-BERT	0.92	0.92	0.92
SVM-Flair	0.87	0.87	0.87
SVM-GloVe	0.83	0.83	0.83
NB-count-vectorizer	0.93	0.93	0.93

As mentioned in section 3.2, we experimented with different pre-trained embeddings for the embedding layer of each deep learning model. Table 4 shows the precision, recall, and F1 scores of the two deep-learning models with different embeddings.For both CNN and LSTM models, GloVe dimension 300 performed significantly better than GloVe dimension The F1 scores for the CNN model are 0.42 and 0.55 for the 100d and 300d GloVe embeddings respectively. For the LSTM

Table 4: Deep learning models results.

Model	Precision	Recall	F1
CNN-GloVe100d	0.77	0.30	0.42
CNN-GloVe300d	0.82	0.42	0.55
CNN-BERT	0.52	0.52	0.52
LSTM-GloVe100d	0.85	0.80	0.83
LSTM-GloVe300d	0.94	0.92	0.93
LSTM-BERT	0.51	0.38	0.41

model, the F1 scores are 0.83 and 0.93 for the 100d and 300d GloVe embeddings respectively. For each of the best performing models in table 4, which are the CNN-GloVe300d and LSTM-GloVe300d, we tune the epoch number, batch size, activation function, and optimization function. Epoch is the number that is used to separate the training into different phases. The best results were produced by using 15 epochs for both models. Batch size is the number that the training data will be divided by. We experimented in the range from 10 to 35, to find that the best batch sizes for the CNN model and the LSTM model were 10 and 25 respectively. Softplus and Softmax activation functions produced the best results for the CNN model and LSTM model respectively. As for the optimization functions, RMSProp and Adam performed best for the CNN model and LSTM model respectively. The results in table 5 were yielded by tuning all the hyperparameters as discussed in this section.

Table 5: Deep learning models results after tuning.

Model	Precision	Recall	F1
CNN-GloVe300d	0.98	0.93	0.95
LSTM-GloVe300d	0.97	0.97	0.97

5 DISCUSSION

We presume that the performance of the machine learning models was comparable to the performance of the deep learning models for our particular task due to the relatively small size of our dataset. A difference in performance is expected to be noticeable if the size of the dataset was larger than it currently is. As well as the number of cognitive distortions. We hypothesize that due to the relatively small size of the dataset, as well as the common structures and keywords between sentences expressing a cognitive distortion, it was easy for the machine learning algorithms to build a distinction between verbal examples of cognitive distortions. We also attribute the similarity in performance between the logistic regression model and the deep learning models to the similarity between the algorithms. (Dreiseitl and Ohno-Machado, 2002) found that logistic regression and neural networks perform on the same level for the majority of the 72 papers that were analyzed. Deep learning can be used to estimate many more parameters on a larger number of permutations than traditional machine learning algorithms. To be able to gain such an advantage, a good ratio between data entries and parameters is required. That's why given a larger dataset with more cognitive distortions than what is currently available, will allow deep learning models to have deeper structures, and to show distinction in results from machine learning algorithms (Young, 2017). Saving the model and loading it into the UI is a simple procedure. Which makes the tool easy to update.

6 CONCLUSIONS

Cognitive distortions put people at risk of developing and sustaining serious mental illnesses. Maintaining unhelpful and negative assumptions affects the overall quality of life. Over time, this sequence among thoughts, emotions, and behaviors can cause or maintain symptoms of depression. Cognitive-behavioral therapy techniques are aimed at recognizing and correcting the patient's misconceptions and maladaptive core beliefs. Our tool can be used to help therapists pay attention to the existence of distorted thoughts that the client has to direct treatment options. It's important to maintain assessments over the course of the treatment, as it can provide the therapists with information about whether the treatment is effective, and to identify if the patient starts developing other cognitive distortions. This tool can be integrated into the assessment and treatment courses seamlessly without any extra steps. Due to the fact that patients already engage in verbal behavior, whether that is verbal communication with the therapist, or through journals. Another useful aspect of this tool is that the patient's verbal behaviors can be monitored through their journals, not just during the therapy session.

In this study, we report the application of machine learning and deep learning techniques toward detecting and classifying cognitive distortions in journaling text. Currently, there is a significant lack of annotated datasets in this domain. Therefore one of our main contributions is the collection and annotation of a novel dataset. We then trained multiple word embeddings and generated a variety of distributed representations of sentences. Which were used to train different machine learning and deep learning algorithms, in order to produce the best performing model. Finally, we developed a user-friendly UI in which the model is integrated.

The lack of access to an annotated dataset formed a setback for this research, in addition to the scarcity of resources for the collection of mental health journals. The tool is targeted for detecting and classifying cognitive distortions in journaling texts, so having a dataset that is collected from real-life mental health journals would improve the accuracy of the tool. Due to the shortage of time and resources, we decided to initiate the study with only two common cognitive distortions. Which makes this study the starting point to an all-inclusive tool for the detection and classification of cognitive distortions. Areas of future investigation definitely include the collection and annotation of a larger dataset, which would improve the accuracy of the classification.

REFERENCES

Asai, A., Evensen, S., Golshan, B., Halevy, A., Li, V., Lopatenko, A., Stepanov, D., Suhara, Y., Tan, W.-c., and Xu, Y. (2018). Happydb: A corpus of 100,000 crowdsourced happy moments.

Asghar, D. M., Subhan, F., Imran, M., Kundi, F., Khan, A., Mosavi, A., Csiba, P., and Varkonyi-Koczy, A. (2020). Performance evaluation of supervised machine learning techniques for efficient detection of emotions from online content. *Computers, Materials and Continua*, 63.

Balabantaray, R. C., Bhubaneswar, I., Mohammad, M., and Sharma, N. (2012). N.: Multi-class twitter emotion classification: A new approach. *International Journal of Applied Information Systems*, pages 48–53.

Baziotis, C., Athanasiou, N., Chronopoulou, A., Kolovou, A., Paraskevopoulos, G., Ellinas, N., Narayanan, S., and Potamianos, A. (2018). Ntua-slp at semeval-2018 task 1: Predicting affective content in tweets with deep attentive rnns and transfer learning.

Beck, A. T. (1970). Cognitive therapy: Nature and relation to behavior therapy. *Behavior Therapy*, 1(2):184–200.

Beck, A. T. (1976). *Cognitive therapy and the emotional disorders*.

Beck, J. S. (2011). *Cognitive behavior therapy: Basics and beyond (2nd ed.)*.

Beck, A. T, R. A. J. S. B. F. . E. G. (1987). *Cognitive therapy of depression*.

Bracewell, D., Minato, J., Ren, F., and Kuroiwa, S. (2006). Determining the emotion of news articles. pages 918– 923.

Bruijniks, S., Bosmans, J., Peeters, F., Hollon, S., Oppen, P., van den Boogaard, T. M., Dingemanse, P., Cuijpers, P., Arntz, A., Franx, G., and Huibers, M. (2015). Frequency and change mechanisms of psychotherapy among depressed patients: Study protocol for a multicenter randomized trial comparing twice-weekly versus once-weekly sessions of cbt and ipt. *BMC psychiatry*, 15:137.

Cachola, I., Holgate, E., Preotiuc-Pietro, D., and Li, J. J. (2018). Expressively vulgar: The socio-dynamics of vulgarity and its effects on sentiment analysis in social media. In *COLING*.

Campan, A., Atnafu, T., Truta, T., and Nolan, J. (2018). Is data collection through twitter streaming api useful for academic research? pages 3638–3643.

Chatterjee, A., Gupta, U., Chinnakotla, M. K., Srikanth, R., Galley, M., and Agrawal, P. (2019). Understanding emotions in text using deep learning and big data. *Computers in Human Behavior*, 93:309–317.

Cliche, M. (2017). Bbtwtr at semeval-2017 task 4: Twitter sentiment analysis with cnns and lstms. pages 573–580.

de Oliveira, I. (2012). *Assessing and Restructuring Dysfunctional Cognitions*.

Dreiseitl, S. and Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. *Journal of Biomedical Informatics*, 35(5):352–359.

Estival, D., Nowak, C., and Zschorn, A. (2004). Towards ontology-based natural language processing. *Proceedings of NLP-XML 2004*.

Gruber, T. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5:199– 220.

Haggag, M., Fathy, S., and Elhaggar, N. (2015). Ontologybased textual emotion detection. *International Journal of Advanced Computer Science and Applications*, 6.

Hasan, M., Agu, E., and Rundensteiner, E. A. (2014). Using hashtags as labels for supervised learning of emotions in twitter messages.

Hofmann, S., Asnaani, A., Vonk, I., Sawyer, A., and Fang, A. (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive therapy and research*, 36:427–440.

Hu, H., Phan, N. H., Geller, J., Iezzi, S., Vo, H., Dou, D., and Chun, S. (2019). An ensemble deep learning model for drug abuse detection in sparse twittersphere.

Minu, R. I. and R.Ezhilarasi (2012). Automatic emotion recognition and classification. volume 38.

Morris, R., Schueller, S., and Picard, R. (2015). Efficacy of a web-based, crowdsourced peer-to-peer cognitive reappraisal platform for depression: Randomized controlled trial. *Journal of Medical Internet Research*, 17:e72.

Mozeticˇ, I., Grcˇar, M., and Smailovic´, J. (2016). Multilingual twitter sentiment classification: The role of human annotators.

Rojas-Barahona, L., Tseng, B.-H., Dai, Y., Mansfield, C., Ramadan, O., Ultes, S., Crawford, M., and Gasˇic´, M. (2018). Deep learning for language understanding of mental health concepts derived from cognitive behavioural therapy. pages 44–54.

Rosenthal, S., Farra, N., and Nakov, P. (2017). Semeval2017 task 4: Sentiment analysis in twitter. pages 502–518.

Ruwaard, J., Lange, A., Schrieken, B., Dolan, C., and Emmelkamp, P. (2012). The effectiveness of online cognitive behavioral treatment in routine clinical practice. *PLoS ONE*, 7.

Shickel, B., Heesacker, M., Benton, S., Ebadi, A., Nickerson, P., and Rashidi, P. (2016). Self-reflective sentiment analysis. pages 23–32.

Shickel, B., Siegel, S., Heesacker, M., Benton, S., and Rashidi, P. (2019). Automatic detection and classification of cognitive distortions in mental health text.

Shivhare, S. N. and Khethawat, S. (2012). Emotion detection from text. volume 2.

Teng, Z., Ren, F., and S, K. (2006). Retracted: Recognition of emotion with svms.

Yamada, D. B., Bernardi, F. A., Miyoshi, N. S. B., de Lima,

I. B., Vinci, A. L. T., Yoshiura, V. T., and Alves, D. (2020). Ontology-based inference for supporting clinical decisions in mental health. In *Computational Science – ICCS 2020*, pages 363–375. Springer International Publishing.

Young, D. (2017). Logistic regression vs deep neural networks [linkedin page].

Yurica, C. and DiTomasso, R. (2005). *Cognitive Distortions*, pages 117–122.