# Using Deep Neural Network to Identify Cancer Survivors Living with Post-Traumatic Stress Disorder on Social Media

Nur Hafieza Ismail[1], Ninghao Liu[1], Mengnan Du[1], Zhe He[2], and Xia Hu[1]

[1] Texas A&M University, College Station, Texas, USA
{refieza,dumengnan,nhliu43,xiahu}@tamu.edu
[2] Florida State University, Tallahassee, Florida, USA
{Zhe.He}@cci.fsu.edu

**Abstract.** Survivors of cancer are at-risk for the lifelong effects of disease and treatment. A significant number of them face Post-Traumatic Stress Disorder (PTSD) that may adversely affect their mental health. Twitter is a social networking site that allows users to interact with others by posting short messages (tweets). These tweets, which to a certain extent reflect the users psychological state, are convenient for data collection. However, Twitter also contains a mix of noisy and genuine tweets. The process of manually identifying the genuine tweets is expensive and time-consuming. Thus, we stream the data using cancer as a keyword and filter the tweets with cancer-free and PTSD related keywords without having to label each tweet manually. Convolutional Neural Network (CNN) learns the representations of the input to identify cancer survivors with PTSD. The experiments on real-world datasets show that the model outperforms the baselines and correctly classifies the new tweets.

**Keywords:** Post-traumatic stress disorder · cancer survivor · social media · deep neural network.

## 1 Background

PTSD is an anxiety disorder that occurs in some people after experiencing or witnessing life-threatening events, may severely affect daily life activities. Being diagnosed with cancer often causes psychological distress due to painful treatments and traumatic experiences of cancer survival [1]. The traditional psychological health diagnosis procedure takes a lot of time and energy which require several interviews, questionnaires, physical evaluation, and testimonies from the caregiver. Twitter is a social media that has simple features that allow users to share their daily feelings [2]. These postings could provide insights on psychological impacts of significant incidents on the users.

Recent work in psychology aims to analyze the manner of self-declared mentally ill users from their interactions and behavior based on written posts [3].

Most of the mental health studies focus only on capturing the psychological problem in society. Besides, the applied analysis approaches such as manual labeling, crowd-sourcing, Twitter Firehouse, and Linguistic Inquiry Word Count (LIWC) are time-consuming and expensive. Thus, data preparation and analysis procedure are challenging. To tackle these challenges, we propose an algorithm to create a labeled dataset using cancer-free related keywords and PTSD features. Also, without having to manually check whether a tweet was written by a cancer survivor with PTSD or not, DNN is able to captures important features from the input dataset. In this work, we used the DNN approach that learns to extract a different level of meaningful representations of texts.

Experts from various fields strive to propose reliable detection models for mental health problems. Despite having the same mission, they touched the issues from different angles. Human emotions can be expressed in many forms of physiological states such as nervous system responses, blood flow, facial expression, and vocal acoustics [4]. However, obtaining information using these ways is usually time-consuming and labor-intensive. The alternative technique for data collection is text postings on social media that are easy, expeditious, and unlimited access to a broader population. A study has shown that Twitter has broad applicability for public health research and can generate valuable knowledge of linguistic style from tweets [2].

Treated cancer patients may find themselves at risk of getting cancer recurrence [5]. During recurrence, the patients reported that treatment decisions are more difficult to make, because the side effects from treatment are more severe, and the fears of uncontrollable pain are greater [6]. This psychological impact that can cause PTSD problem is one of the significant concerns in clinical oncology [7]. Receiving immediate attention to PTSD can help to improve the quality of life more quickly. Various analysis methods such as supervised and unsupervised machine learning models have been adopted for the detection and monitoring of PTSD. In this work, we developed a model using DNN approach that able to learn from different levels of representation of text input.

## 2 The Proposed Framework for Cancer Survivors Living with PTSD Diagnosis on Social Media

Figure 1 illustrates the proposed framework for identifying cancer survivors living with PTSD on social media. First, we extract patterns about the particular words used by sufferers from conventional studies on depression. Second, the extracted patterns are then used to identify tweets that contain PTSD symptoms on the cancer survivors dataset. The detailed process of our proposed framework will be explained in three subsections below.

*Feature Extraction:* The top part of Figure 1 shows the overview of previous work related to depression detection using social media data [8]. It employs the crowd-sourcing approach to identify tweets associated with mental illness, which are then labeled as PTSD positive dataset. Tweets that are not identified to

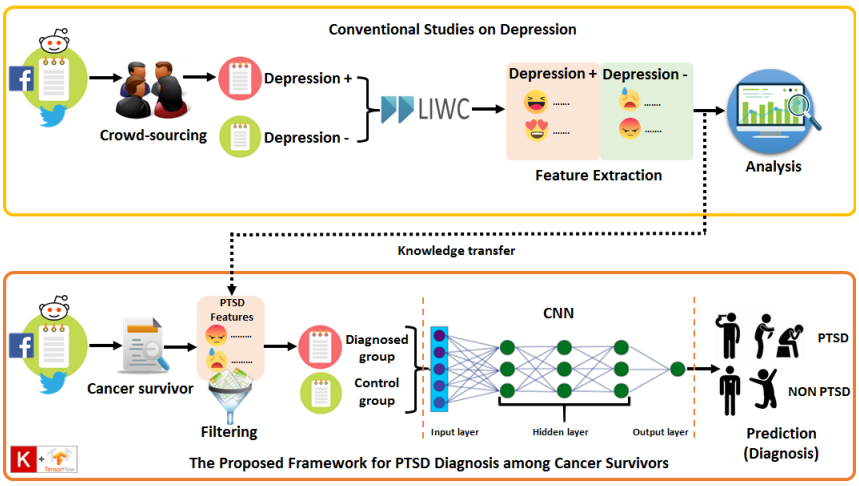


**Fig. 1.** The overview of the proposed work in identifying PTSD.

be related to mental illness are labeled as PTSD negative dataset. The process continues by combining PTSD positive and negative datasets. The goal of this process is to understand the differences of linguistic style of both groups.

*Knowledge Transfer:* In this work, knowledge transfer can be defined as a process that uses depression feature extraction outputs to develop a labeled PTSD dataset. Most PTSD sufferers also have depression in diverse epidemiological samples. This comorbidity reflects overlapping symptoms in both disorders [9]. Thus, we opted to use depression lexicon taken from [8] for PTSD tweets identification for this work. The lower part of Figure 1 presents our proposed framework for identifying of cancer survivors living with PTSD. We collected the raw dataset using 'cancer' as a keyword through the Twitter's Application Programming Interface (API) in a period of four months from October 2017 - January 2018. Then, we conducted the extraction process in two steps using two sets of keywords (cancer survivor and depression) to create a ground truth dataset. This process is called as 'knowledge transfer' in which published information is taken as a guide for our proposed model.

*CNN Architecture:* We adopted only one convolutional layer with embedding layer in the CNN network setting to produce results for tweets classification. It requires the specification of the vocabulary size, the size of the real-valued vector space, and the maximum length of words in input tweets. For convolutional feature maps, we used word embedding with 100-dimensional for text representation. 32 filters were applied with a kernel size of 8 and a rectified linear (ReLu) activation function. Followed by a pooling layer, the filters will generate feature maps and reduces the output by half. The end layer uses a sigmoid activation function to output a value between two categories of positive and negative in the tweets based on the concatenated the previous vectors.

## 3 Experiment

In this experiments, the dataset with PTSD positive represents the diagnosed group, while PTSD negative represents the control group. For the diagnosed group, we retrieved tweets from a user who publicly stated that they survived cancer and had PTSD. To construct the PTSD negative group, we make uses of tweets from 'Twitter User Gender Classification' dataset from Kaggle website [3]. Both groups have the same total tweets of 10k to create balanced datasets. The data preparation phase includes three tasks: (1) splitting the dataset into 80% for training and 20% for testing, (2) cleaning the dataset to remove punctuation, stop words, and numbers; (3) defining a vocabulary of preferred words from a training dataset by stepping through words and keeping only tokens with minimum occurrences of five. We used Keras API running on Tensorflow to train deep learning models. All the models were trained with ten epochs through the training data. The efficient Adam implementation of stochastic gradient descent was used.

The three baselines that are capable of handling text dataset used for evaluating our proposed algorithm are Multiple Layer Perceptron (MLP), CNN n-gram, and Recurrent Neural Network (RNN). Our results indicate that CNN can effectively identify cancer survivor with PTSD. Experimental results in Table 1 show the accuracy of CNN of 98.5% is higher than CNN n-gram by 0.2%. We ran the experiments multiple times due to the stochastic nature of DNN to get the reasonably accurate result.

| Method | Accuracy (%) |
|---|---|
| CNN | 98.5 |
| MLP | 95.5 |
| CNN N-gram | 98.3 |
| RNN | 96.9 |

**Table 1.** Experiment results of identifying cancer survivors with PTSD.

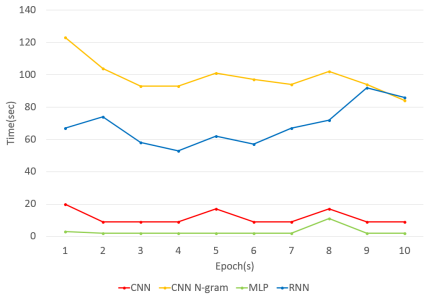


**Fig. 2.** The learning time taken.

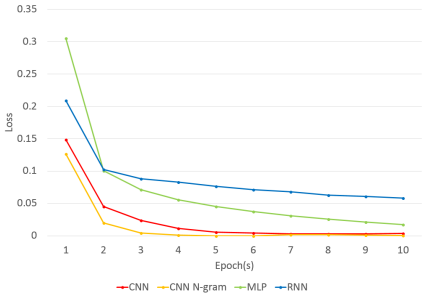


**Fig. 3.** The loss values.

[3] https://www.kaggle.com/crowdflower/twitter-user-gender-classificationgender-classifier-DFE-791531.csv

Figure 2 presents the time taken during the training process with CNN and MLP, which took less time compared to CNN n-gram and RNN. Even though CNN and CNN n-gram yield almost the same accuracy, the training time taken by CNN n-gram is 14.5 minutes longer than CNN. Figure 3 shows the loss values in training set of all models, CNN and CNN n-gram display low losses.

## 4 Discussion and Conclusion

PTSD is a severe anxiety disorder that affects individuals who are exposed to traumatic events such as cancer. Cancer survivors are at risk of short-term or long-term effects on physical and psychosocial well being. Therefore, the evaluation and treatment of PTSD are essential parts of cancer survivorship care. We propose a prediction system with a CNN model that can produce promising results. Experimental results demonstrated that CNN was able to capture important signals from texts in determining PTSD among cancer survivors. The social media users with cancer history who suffer from depression will benefit from the prediction system. It will act as an alarming system by detecting the depression presence based on users' postings.

## References

1. Alkan, A., Guc, Z. G., Senler, F. C., Yavuzsen, T., Onur, H., Dogan, M., ... & Turhal, S. (2016). Breast cancer survivors suffer from persistent postmastectomy pain syndrome and posttraumatic stress disorder (ORTHUS study): a study of the palliative care working committee of the Turkish Oncology Group (TOG). Supportive Care in Cancer, 24(9), 3747-3755.
2. Paul, M. J., & Dredze, M. (2011). You are what you Tweet: Analyzing Twitter for public health. Icwsm, 20, 265-272.
3. Harman, G. A. C. C. T., & Dredze, M. H. (2014). Measuring post traumatic stress disorder in Twitter. In ICWSM.
4. Krothapalli, S. R., & Koolagudi, S. G. (2013). Characterization and recognition of emotions from speech using excitation source information. International journal of speech technology, 16(2), 181-201.
5. Fischer, U., Zachariae, O., Baum, F., von Heyden, D., Funke, M., & Liersch, T. (2004). The influence of preoperative MRI of the breasts on recurrence rate in patients with breast cancer. European radiology, 14(10), 1725-1731.
6. Lewis, F. M., & Deal, L. W. (1995, July). Balancing our lives: A study of the married couple's experience with breast cancer recurrence. In Oncology Nursing Forum-Oncology Nursing Society (Vol. 22, No. 6, pp. 943-956). [Pittsburgh, PA, etc.] Oncology Nursing Society.
7. Stuber, M. L., Kazak, A. E., Meeske, K., Barakat, L., Guthrie, D., Garnier, H., ... & Meadows, A. (1997). Predictors of posttraumatic stress symptoms in childhood cancer survivors. Pediatrics, 100(6), 958-964.
8. De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting depression via social media. ICWSM, 13, 1-10.
9. Flory, J. D., & Yehuda, R. (2015). Comorbidity between post-traumatic stress disorder and major depressive disor-der: alternative explanations and treatment considerations. Dialogues in clinical neuroscience, 17(2), 141.