

Multi-Task Learning to Capture Changes in Mood Over Time

Prasadith Buddhitha, Ahmed Hussein Orabi, Mahmoud Hussein Orabi, Diana Inkpen

School of Electrical Engineering and Computer Science

University of Ottawa, Ottawa, ON, K1N 6N5, Canada

{pgamaara, ahuss045, mhuss092, diana.inkpen}@uottawa.ca

Abstract

This paper investigates the impact of using Multi-Task Learning (MTL) to predict mood changes over time for each individual (social media user). The presented models were developed as a part of the Computational Linguistics and Clinical Psychology (CLPsych) 2022 shared task. Given the limited number of Reddit social media users, as well as their posts, we decided to experiment with different multi-task learning architectures to identify to what extent knowledge can be shared among similar tasks. Due to class imbalance at both post and user levels and to accommodate task alignment, we randomly sampled an equal number of instances from the respective classes and performed ensemble learning to reduce prediction variance. Faced with several constraints, we managed to produce competitive results that could provide insights into the use of multi-task learning to identify mood changes over time and suicide ideation risk.

1 Introduction

For many countries, suicide has been a formidable challenge, where 1.3% of world deaths in 2019 were due to suicides. Of the committed suicides, most of them were by individuals before reaching their fifties and in countries with low to middle income (World Health Organization, 2021). Considering these factors, it is of utmost importance for any institution responsible for the mental health of the population to early detect users susceptible to suicide ideation and mental disorders. In recent years, social media has become an integral part of the everyday life of many. According to Schimmele et al. (2021), more than 25% of users aged between 15 and 64 have shared their personal information (e.g., pictures, videos, text-based posts) publicly. This data rich with personal information opens the pathway for many research opportunities.

The importance of using social media data to detect users susceptible to suicide ideation (MacA-

vane et al., 2021; Zirikly et al., 2019) and mental disorders (Coppersmith et al., 2015b; Milne et al., 2016) was demonstrated throughout the CLPsych workshop series. When analyzing research, including publications in the CLPsych workshop series, we could see that in comparison to traditional machine learning methods (Cohan et al., 2016; Coppersmith et al., 2015a; Jamil et al., 2017; Schwartz et al., 2014), recent research has focused more on using deep learning architectures (Hussein Orabi et al., 2018; Kshirsagar et al., 2017; Mohammadi et al., 2019) that considerably reduce the time and effort required for feature engineering. However, researchers have continued using traditional machine learning methods to predict individuals susceptible to mental disorders and suicide ideation, which could be due to the lack of large sets of annotated data (e.g., Hauser et al. (2019) or to the requirement of explainability (e.g., Saha et al. (2022)).

In this paper, we describe the experiments conducted using deep learning methods, specifically with multi-task learning, to predict a user's mood change over time (i.e., either a switch or an escalation in the mood) and also the suicide ideation risk level where a selected user can be categorized into one of the following risk categories: low, moderate, or severe. The main reason for selecting multi-task learning is to leverage its capabilities of sharing knowledge between related but different tasks that could potentially alleviate the negative impact of having a small number of training instances. For example, we identified the negative impact of having a limited number of data points during model training, specifically when using deep learning architectures where different regularization methods were used to reduce model overfitting and increase the model's generalizability. When predicting suicide ideation risk level, we used an additional dataset from Cohan et al. (2018), named the Self-Reported Mental Health Diagnoses (SMHD) dataset, which

consists of users who have self-declared mental disorders. Similar to [Gamaarachchige \(2021\)](#), which demonstrates the impact different mental disorders have on suicide ideation detection (i.e., whether an individual is susceptible to suicide ideation or not), we investigate the impact mental disorders have on different suicide ideation risk levels (i.e., low, moderate, or severe).

2 Task and Data

The CLPsych 2022 shared task consisted of two subtasks ([Tsakalidis et al., 2022a](#)). The first task was to identify a user’s mood change over time ([Tsakalidis et al., 2022b](#)), and the second task was to predict the level of suicidality risk for an individual ([Shing et al., 2018](#); [Zirikly et al., 2019](#)). Then, when predicting the suicidality risk, the participants were encouraged to discover if there is any relationship between the mood change over time and the risk of suicidality. The dataset provided to the task participants consisted of users and their posts extracted from the Reddit social media platform. Apart from 3,089 posts distributed across 139 timelines posted by 83 users, the rest of the users were sampled from the University of Maryland Reddit Suicidality Dataset ([Shing et al., 2018](#); [Zirikly et al., 2019](#)) and the eRisk dataset ([Losada and Crestani, 2016](#); [Losada et al., 2020](#)). The combined dataset statistics are shown in table 1.

# Timelines	Users	Posts
204	149	5,063

Table 1: CLPsych 2022 training data.

For both tasks, we combined the text fields “title” and “content”, and after several preliminary preprocessing steps, we identified 5,143 posts where the majority of the posts were categorized as “None”. The distribution of the classes in the training dataset is shown in table 2, for Task A.

Label	Count	Percentage
None (O)	4,043	79%
Escalation (IE)	773	15%
Switch (IS)	327	6%

Table 2: Post-level class distribution.

For "Task B", we grouped all the posts per user and trained our proposed deep learning model on a

dataset that contained 127 users distributed among three classes as shown in table 3.

Label (risk level)	Count	Percentage
Low	11	9%
Moderate	55	43%
Severe	61	48%

Table 3: Suicide ideation risk level class distribution.

A considerable class imbalance can be identified when analyzing the class distribution for both tasks. Such imbalance could adversely impact model training and its generalizability, which we will discuss more in the following sections.

For “Task B” only, we used an external dataset from [Cohan et al. \(2018\)](#), that contains users who have self-declared single or multiple mental disorders. Based on the conclusions derived by [Gamaarachchige \(2021\)](#), we sampled users who have self-declared Post-Traumatic Stress Disorder (PTSD), Anxiety, and Bipolar Disorder as the input for the mental illness detection task within the MTL environment. However, we did not include any users who have self-declared other mental illnesses due to time constraints.

Macro-averaged precision, recall, and F1-score were used as evaluation metrics at the post, timeline, and coverage levels.

To generalize and reduce input noise, we performed the following preprocessing steps: lower-cased the texts, kept only a selected set of stop words, removed most of the non-alphanumeric characters, removed numbers and URLs, and expanded contractions.

3 Methodology

As mentioned before, we based our experiments on multi-task learning and specifically an architecture using a combination of soft and hard parameter sharing. Multi-task learning allows related tasks to share representations ([Caruana, 1997](#)), and based on how parameters are being shared, can be categorized into two types of architectures, which are hard parameter sharing and soft parameter sharing ([Ruder, 2017](#)). Each task will share model weights in hard parameter sharing, and features unique to individual tasks will be extracted through the task-specific layers. Even though model weights are not shared between layers in soft parameter sharing, the parameters are regularized between the layers to discover similarities. We used a custom loss

function that combines "categorical cross-entropy", "mean squared error", and "cosine similarity" to regularize layer weights.

When using MTL for "Task A" (i.e., according to figure 1), the two tasks were to predict whether the post is a "Switch" or "None" (i.e., "IS" or "O") or whether it is an "Escalation" or "None" (i.e., "IE" or "O"). To prepare the training and validation input for each task, we sampled an equal number of instances from each class where the number of instances to sample is based on the minority class. Selecting an equal number of instances for each class made it possible to align the tasks so that similar tasks could potentially share a common feature space. We kept aside a sample with a class distribution to be the same as the original dataset for testing.

For "Task A", the task-specific layers consist of a multi-channel Convolutional Neural Network (CNN) (Kim, 2014) where each channel was responsible for filtering features constituting bigrams and trigrams. To reduce the number of learnable parameters, the output from the CNN layers was further transformed using Global Maximum Pooling and then sent through a feedforward neural network. The output from each channel was then merged to form vectors that represent the task-specific features. These vectors were submitted to a loss function to regularize the network weights further. The merged outputs from each task-specific layer were concatenated to form the shared representation where each task will learn from a common feature space. It was identified that the model started to overfit the training data within a few epochs and consequently generated poor results during inference. To overcome model overfitting, we used several regularization techniques such as dropout (Srivastava et al., 2014) (i.e., a probability of 0.4 for "Task A" and 0.2 for "Task B") and L1 and L2 regularization to penalize larger weights in the multi-channel CNN. Further experiments discovered that making the model more or less complex reduced prediction accuracies due to either overfitting or underfitting, respectively.

We adopted an ensemble learning approach to reduce the variance in the results, which could be due to noise and random sampling. Model training and evaluation were done on three stratified training and validation splits where the final output is generated using an ensemble strategy on the combined predictions. We used the model averaging

ensemble (Brownlee, 2018) strategy to generate the output.

For "Task B", we used the same methods as for "Task A", except that we used an additional dataset to enhance the shared feature space between users susceptible to suicide ideation and mental disorders. Therefore, we selected a random sample of users similar to the number of users in the suicide ideation detection dataset. For example, to extract shared hidden features between users with severe suicide ideation risk and PTSD, we randomly selected 61 users who have self-declared PTSD from the SMHD dataset. The number of users is identified from the training dataset, where 61 users are categorized with severe suicide ideation risk.

The output of the suicide ideation detection task predicts three classes, that is, whether the user has a "Low", "Moderate", or "Severe" suicide ideation risk. For the second task, we conducted experiments using a different combination of mental disorders by predicting whether a given user has PTSD, Anxiety, or Bipolar Disorder. The final predictions are based on a model where users with "Moderate" and "Severe" suicide risks were aligned (i.e., sharing a common feature space) with users who have self-declared PTSD, and users with "Low" risk were aligned with users who have self-declared anxiety.

We used randomly initialized and trainable embedding layers with a dimension of 300 units for both subtasks. For task-specific layers, we used Rectified Linear Unit (ReLU) (2010) activation function and Adam optimizer (2015) with a learning rate 0.001 to update network weights.

4 Experiments and Results

We trained our models for fifteen epochs and reduced the learning rate by a factor of 0.1 if the validation loss did not improve. If the validation loss did not continuously improve, we stopped training and returned the model weights that produced the minimum loss. For both tasks, we trained our models using a mini-batch of size 16. Finally, we selected the label with the highest probability from the output generated using the model averaging ensemble.

We submitted three results for "Task A" and one for "Task B". The difference between our two submissions, "uOttawa-AI(2)" and "uOttawa-AI(3)", is based on regularization, where with more optimized regularization hyperparameters (i.e., on the

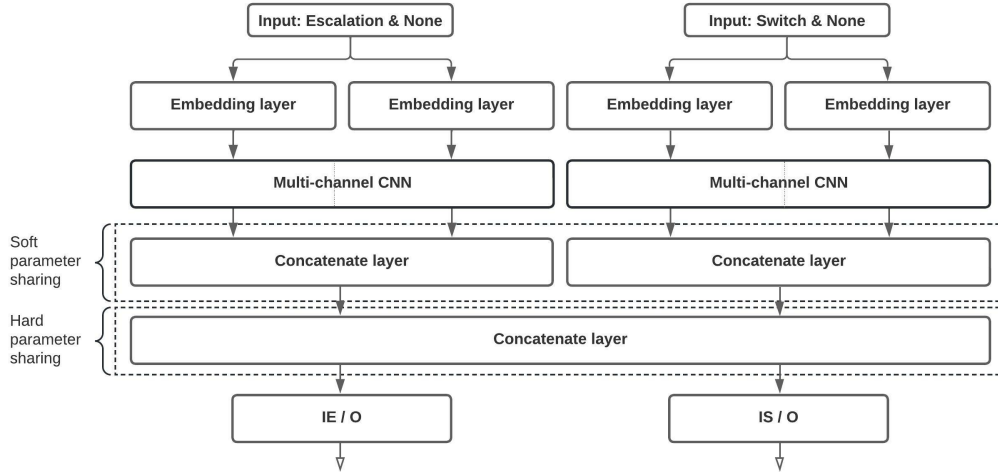


Figure 1: The Proposed multi-task learning architecture with hard and soft parameter sharing. The mentioned architecture is used mainly for "Task A". For "Task B", instead of IE/O and IS/O, we use suicide ideation risk levels as one output and the selected mental disorders as the second (i.e., PTSD/Anxiety).

submission uOttawa-AI(2)), we managed to train our model for more epochs and as a result produced a more generalized model. The “uOttawa-AI(1)” submission results are from a model with fewer learnable parameters.

Our results, compared to a majority class baseline and two preliminary experiments conducted by the task organizers (Tsakalidis et al., 2022b), are mentioned in tables 4, 5, 6, and 7. The results are macro averaged at the post level, window-based, and coverage-based (please refer Tsakalidis et al. (2022a) for more details on the evaluation metrics).

	Precision	Recall	F1
uOttawa-AI(2)	0.504	0.529	0.511
Majority	nan	0.333	0.280
TFIDF	0.545	0.495	0.492
BERT	0.522	0.386	0.380

Table 4: Post-level macro averaged results.

	Precision	Recall
uOttawa-AI(2)	0.347	0.453
Majority	nan	0.141
TFIDF	0.377	0.424
BERT	0.260	0.204

Table 5: Coverage-based macro averaged results.

5 Discussion

When analyzing the results of “Task A”, we could see that our proposed architecture has produced competitive results when compared against the baseline and two of the preliminary experiments that use TF-IDF features with logistic regression and the BERT (Bidirectional Encoder Representations from Transformers) language model trained using the Talklife dataset (Tsakalidis et al., 2022b). We also identified that our submission “uOttawa-AI(2)” has produced better coverage and window-based (refer to table 6) predictions.

Even though the test results for the “Task B” model have produced better outcomes than the majority class baseline and the preliminary models trained by the task organizers (refer to table 7, our model has not performed well in comparison to the best results. One of the critical reasons for the low results is class imbalance. During training, there were only 11 instances for the “Low” risk class compared to 55 and 61 for “Moderate” and “Severe” risk (refer to table 3). During inference, our model has not predicted “Low” risk labels but only “Moderate” and “Severe” labels. Another reason that we identified is the use of mental illness data as a complementary task. Even though the mental illness detection task has shared a common feature space with the suicide ideation detection task (i.e., suicide ideation or not) in Gamaarachchige (2021), when it comes to a more granular level (i.e., level of risk), mental ill-

	Window 1		Window 2		Window 3	
	P	R	P	R	P	R
uOttawa-AI(2)	0.529	0.621	0.559	0.662	0.596	0.691
Majority	nan	0.333	nan	0.333	nan	0.333
TFIDF	0.496	0.539	0.505	0.550	0.506	0.551
BERT	0.582	0.392	0.608	0.405	0.608	0.405

Table 6: Window-based macro averaged results.

	Precision	Recall	F1
uOttawa-AI	0.329	0.365	0.344
Majority	0.156	0.333	0.212
TFIDF	0.302	0.338	0.295

Table 7: Task B macro averaged results.

ness detection task has not managed to share features with suicide ideation risk levels. Even though we could not derive a conclusion on the suicide risk level and its correlation with a particular mental disorder, it could be assumed that more data representing different risk categories could derive a stronger relationship with certain mental disorders.

6 Conclusion and Future Work

We have investigated the applicability of multi-task learning to predict the change in mood of a social media user over time. With limited experiments, we managed to identify that MTL can be effectively applied to predict whether a post contains a mood shift, an escalation, or no change. Using different MTL architectures, which adopted different forms of parameter sharing strategies, it was identified that a combination of both the parameter sharing strategies (i.e., hard and soft parameter sharing) managed to produce better results. The main drawbacks we faced when using deep learning methods for classification are the class imbalance and the limited number of data points. For both tasks, we adopted a sampling strategy that facilitates task alignment. For “Task B”, we introduced a complementary task intending to enrich the hidden features space so that we could, to a certain extent, eliminate the negative impact of having a smaller dataset with class imbalance. When analyzing the prediction outcomes, we could assume that features shared by certain mental disorders are not sufficient to define a decision boundary over suicide ideation risk levels.

In future research, we will look into the possibilities of improving the prediction accuracies

by making changes to the current architecture (e.g., by changing the constructs of task-specific and shared layers) and also by adding contextual (e.g., ELMo¹ (Peters et al., 2018), BERT (Devlin et al., 2019)) and non-contextual embeddings (e.g., word2vec (Mikolov et al., 2013), fastText (Joulin et al., 2017)).

Ethics Statement

Secure access to the shared task dataset was provided with IRB approval under the University of Maryland, College Park protocol 1642625 and approval by the Biomedical and Scientific Research Ethics Committee (BSREC) at the University of Warwick (ethical application reference BSREC 40/19-20).

Acknowledgements

The authors are particularly grateful to the anonymous users of Reddit whose data feature in this year’s shared task dataset, to the annotators of the data for Task A, to the clinical experts from Bar-Ilan University who annotated the data for Task B, and the American Association of Suicidology, to NORC for creating and administering the secure infrastructure and providing researcher support and to UKRI for providing funding to the CLPsych 2022 shared task organizers.

We would also like to thank the anonymous reviewers for their in-depth feedback and guidance, and the CLPsych shared task organizers for organizing the task and especially Adam Tsakalidis for continuously supporting us through the challenging times. Finally, we would like to convey our appreciation to all who have made it possible for the research domain to move forward by collecting, annotating and distributing data and conducting research to discover valuable insights.

¹Embeddings from Language Models

References

- Jason Brownlee. 2018. *Better Deep Learning: Train Faster, Reduce Overfitting, and Make Better Predictions.*, 1 edition. Machine Learning Mastery.
- Rich Caruana. 1997. [Multitask Learning](#). *Machine Learning*, 28(1):41–75.
- Arman Cohan, Bart Desmet, Andrew Yates, Luca Soldaini, Sean MacAvaney, and Nazli Goharian. 2018. SMHD: A Large-Scale Resource for Exploring Online Language Usage for Multiple Mental Health Conditions. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1485–1497, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Arman Cohan, Sydney Young, and Nazli Goharian. 2016. Triaging Mental Health Forum Posts. In *Proceedings of the 3rd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 143–147, San Diego, CA, USA. Association for Computational Linguistics.
- Glen Coppersmith, Mark Dredze, Craig Harman, and Kristy Hollingshead. 2015a. [From ADHD to SAD: Analyzing the Language of Mental Health on Twitter through Self-Reported Diagnoses](#). In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 1–10, Denver, Colorado. Association for Computational Linguistics.
- Glen Coppersmith, Mark Dredze, Craig Harman, Hollingshead Kristy, and Margaret Mitchell. 2015b. [CLPsych 2015 Shared Task: Depression and PTSD on Twitter](#). In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 31–39, Denver, Colorado. Association for Computational Linguistics.
- 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*, pages 4171–4186, Minneapolis, MN, USA. Association for Computational Linguistics.
- Prasadith Buddhitha Kirinde Gamaarachchige. 2021. [Mental Illness and Suicide Ideation Detection Using Social Media Data](#). Ph.D. thesis, University of Ottawa.
- Michael Hauser, Evangelos Sariyanidi, Birkan Tunc, Casey Zampella, Edward Brodtkin, Robert Schultz, and Julia Parish-Morris. 2019. Using natural conversations to classify autism with limited data: Age matters. In *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, pages 45–54, Minneapolis, Minnesota. Association for Computational Linguistics.
- Ahmed Husseini Orabi, Prasadith Buddhitha, Mahmoud Husseini Orabi, and Diana Inkpen. 2018. [Deep Learning for Depression Detection of Twitter Users](#). In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, pages 88–97, New Orleans, LA. Association for Computational Linguistics.
- Zunaira Jamil, Diana Inkpen, Prasadith Buddhitha, and Kenton White. 2017. [Monitoring Tweets for Depression to Detect At-risk Users](#). In *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology*, pages 32–40, Vancouver. Association for Computational Linguistics.
- Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of Tricks for Efficient Text Classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pages 427–431, Valencia, Spain. Association for Computational Linguistics.
- Yoon Kim. 2014. [Convolutional neural networks for sentence classification](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar. Association for Computational Linguistics.
- Diederik P. Kingma and Jimmy Lei Ba. 2015. Adam: A Method for Stochastic Optimization. In *Proceedings of the 3rd International Conference on Learning Representations 2015*, pages 1–15, San Diego, CA, USA.
- Rohan Kshirsagar, Robert Morris, and Samuel Bowman. 2017. [Detecting and Explaining Crisis](#). In *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology — From Linguistic Signal to Clinical Reality*, pages 66–73, Vancouver, BC. Association for Computational Linguistics.
- David E. Losada and Fabio Crestani. 2016. A test collection for research on depression and language use. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction, 7th International Conference of the {CLEF} Association*, volume 9822 LNCS, Evora, Portugal.
- David E. Losada, Fabio Crestani, and Javier Parapar. 2020. [Overview of eRisk 2020: Early Risk Prediction on the Internet](#). *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12260 LNCS:272–287.
- Sean MacAvaney, Anjali Mittu, Glen Coppersmith, Jeff Leintz, and Philip Resnik. 2021. [Community-level Research on Suicidality Prediction in a Secure Environment: Overview of the CLPsych 2021 Shared Task](#). In *Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology*, pages 70–80, Online. Association for Computational Linguistics.

- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems*, page 3111–3119, Red Hook, NY, USA. Curran Associates Inc.
- David N Milne, Glen Pink, Ben Hachey, and Rafael A Calvo. 2016. [CLPsych 2016 Shared Task : Triaging content in online peer-support forums](#). In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, pages 118–127, San Diego, CA, USA. Association for Computational Linguistics.
- Elham Mohammadi, Hessam Amini, and Leila Kosseim. 2019. [CLaC at CLPsych 2019: Fusion of Neural Features and Predicted Class Probabilities for Suicide Risk Assessment Based on Online Posts](#). In *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, pages 34–38. Association for Computational Linguistics.
- Vinod Nair and Geoffrey E Hinton. 2010. Rectified Linear Units Improve Restricted Boltzmann Machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML'10*, page 807–814, Madison, WI, USA. Omnipress.
- Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2227–2237, New Orleans, Louisiana. Association for Computational Linguistics.
- Sebastian Ruder. 2017. An Overview of Multi-Task Learning in Deep Neural Networks. *CoRR*, abs/1706.0.
- Koustuv Saha, Asra Yousuf, Ryan L. Boyd, James W. Pennebaker, and Munmun De Choudhury. 2022. [Social Media Discussions Predict Mental Health Consultations on College Campuses](#). *Scientific Reports*, 12(1):1–11.
- Christoph Schimmele, Jonathan Fonberg, and Grant Schellenberg. 2021. [Canadians’ assessments of social media in their lives](#). Technical report, Statistics Canada.
- H Andrew Schwartz, Johannes Eichstaedt, Margaret L Kern, Gregory Park, Maarten Sap, David Stillwell, Michal Kosinski, and Lyle Ungar. 2014. [Towards Assessing Changes in Degree of Depression through Facebook](#). In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 118–125, Baltimore, Maryland, USA. Association for Computational Linguistics.
- Han-chin Shing, Suraj Nair, Ayah Zirikly, Meir Friedenberg, Hal Daumé III, and Philip Resnik. 2018. [Expert, Crowdsourced, and Machine Assessment of Suicide Risk via Online Postings](#). In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, pages 25–36, New Orleans, LA. Association for Computational Linguistics.
- Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15(2):1929–1958.
- Adam Tsakalidis, Jenny Chim, Iman Munire Bilal, Ayah Zirikly, Dana Atzil-Slonim, Federico Nanni, Philip Resnik, Manas Gaur, Kaushik Roy, Becky Inkster, Jeff Leintz, and Maria Liakata. 2022a. Overview of the CLPsych 2022 Shared Task: Capturing Moments of Change in Longitudinal User Posts. In *Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology: Mental Health in the Face of Change*.
- Adam Tsakalidis, Federico Nanni, Anthony Hills, Jenny Chim, Jiayu Song, and Maria Liakata. 2022b. Identifying Moments of Change from Longitudinal User Text. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, volume 1, pages 4647–4660, Dublin, Ireland. Association for Computational Linguistics.
- World Health Organization. 2021. [Suicide worldwide in 2019: global health estimates](#). Technical report, World Health Organization, Geneva.
- Ayah Zirikly, Philip Resnik, Ozlem Uzuner, and Kristy Hollingshead. 2019. [CLPsych 2019 Shared Task: Predicting the Degree of Suicide Risk in Reddit Posts](#). In *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, pages 24–33, Minneapolis, Minnesota. Association for Computational Linguistics.