

COLING 2020

**Social Media Mining for Health Applications (#SMM4H)
Workshop & Shared Task**

Proceedings of the Fifth Workshop

December 12, 2020
Barcelona, Spain (Online)

Copyright of each paper stays with the respective authors (or their employers).

ISBN 978-1-952148-41-5

Preface

Welcome to the 5th Social Media Mining for Health Applications Workshop & Shared Task (#SMM4H 2020), co-located at the 28th International Conference on Computational Linguistics (COLING 2020). In its fifth iteration, #SMM4H 2020 continues to serve as a venue for bringing together researchers interested in addressing the significant opportunities and challenges of utilizing the vast amount of data on social media for health informatics. For #SMM4H 2020, we accepted 5 workshop papers (acceptance rate of 56%) and 26 shared task system description papers. Each submission was peer-reviewed by two reviewers.

The accepted workshop papers span a range of social media data—Twitter, Facebook, Reddit, and online health forums—and health domains, including diabetes, depression, COVID-19, medical misinformation, and adverse drug reactions. Cornelius et al. present an online platform that aggregates and visualizes methods for extracting information related to COVID-19 on Twitter. Dirkson et al. explore modeling conversational features of posts, in addition to the posts themselves, for detecting adverse drug reactions on Facebook, and medical misinformation. Romberg et al. present an annotated, German-language corpus for extracting information needs expressed online by patients with diabetes. Moßburger et al. use various text mining techniques to compare features of depression forums on Reddit and a curated, moderated site. Finally, Owen et al. present an annotated, English-language corpus for detecting depression and anxiety on Twitter.

The #SMM4H 2020 shared tasks sought to advance the use of Twitter data (tweets) for pharmacovigilance, toxicovigilance, and epidemiology of birth defects. In addition to re-runs of three tasks, #SMM4H 2020 included new tasks for detecting adverse drug reactions in French and Russian tweets, characterizing chatter related to prescription medication abuse, and detecting self reports of birth defect pregnancy outcomes. The five tasks required methods for binary classification, multi-class classification, and named entity recognition (NER). With 29 teams and a total of 130 system submissions, participation in the #SMM4H shared tasks continues to grow. Among the 26 shared task system description papers that were accepted, 6 teams were invited to present their system orally.

The organizing committee of #SMM4H 2020 would like to thank the program committee, the additional reviewers of system description papers, the organizers of COLING 2020 (especially the workshop co-chairs), the annotators of the shared task data, and, of course, everyone who submitted a paper or participated in the shared tasks. #SMM4H 2020 would not have been possible without all of them.

Graciela, Ari, Ivan, Davy, Arjun, Karen, Abeed, Anne-Lyse, Elena, Zulfat, Ilseyar

Organizing and Program Committees

Organizing Committee:

Graciela Gonzalez-Hernandez, University of Pennsylvania, USA
Ari Z. Klein, University of Pennsylvania, USA
Ivan Flores, University of Pennsylvania, USA
Davy Weissenbacher, University of Pennsylvania, USA
Arjun Magge, University of Pennsylvania, USA
Karen O'Connor, University of Pennsylvania, USA
Abeed Sarker, Emory University, USA
Anne-Lyse Minard, Université d'Orléans, France
Elena Tutubalina, Kazan Federal University, Russia
Zulfat Miftahutdinov, Kazan Federal University, Russia
Ilseyar Alimova, Kazan Federal University, Russia

Program Committee:

Robert Leaman, US National Library of Medicine, USA
Diego Molla, Macquarie University, Australia
Azadeh Nikfarjam, Apple, USA
Thierry Poibeau, French National Center for Scientific Research, France
Kirk Roberts, University of Texas Health Science Center at Houston, USA
Yutaka Sasaki, Toyota Technological Institute, Japan
H. Andrew Schwartz, Stony Brook University, USA
Nicolas Turenne, French National Institute for Agricultural Research, France
Karin Verspoor, University of Melbourne, Australia
Pierre Zweigenbaum, French National Center for Scientific Research, France

Table of Contents

<i>COVID-19 Twitter Monitor: Aggregating and Visualizing COVID-19 Related Trends in Social Media</i> Joseph Cornelius, Tilia Ellendorff, Lenz Furrer and Fabio Rinaldi	1
<i>Conversation-Aware Filtering of Online Patient Forum Messages</i> Anne Dirkson, Suzan Verberne and Wessel Kraaij	11
<i>Annotating Patient Information Needs in Online Diabetes Forums</i> Julia Romberg, Jan Dyczmons, Sandra Olivia Borgmann, Jana Sommer, Markus Vomhof, Cecilia Brunoni, Ismael Bruck-Ramisch, Luis Enders, Andrea Icks and Stefan Conrad	19
<i>Overview of the Fifth Social Media Mining for Health Applications (#SMM4H) Shared Tasks at COLING 2020</i> Ari Klein, Ilseyar Alimova, Ivan Flores, Arjun Magge, Zulfat Miftahutdinov, Anne-Lyse Minard, Karen O'Connor, Abeed Sarker, Elena Tutubalina, Davy Weissenbacher and Graciela Gonzalez-Hernandez	27
<i>Ensemble BERT for Classifying Medication-mentioning Tweets</i> Huong Dang, Kahyun Lee, Sam Henry and Özlem Uzuner	37
<i>ISLab System for SMM4H Shared Task 2020</i> Chen-Kai Wang, Hong-Jie Dai, You-Chen Zhang, Bo-Chun Xu, Bo-Hong Wang, You-Ning Xu, Po-Hao Chen and Chung-Hong Lee	42
<i>BERT Implementation for Detecting Adverse Drug Effects Mentions in Russian</i> Andrey Gusev, Anna Kuznetsova, Anna Polyanskaya and Egor Yatsishin	46
<i>KFU NLP Team at SMM4H 2020 Tasks: Cross-lingual Transfer Learning with Pretrained Language Models for Drug Reactions</i> Zulfat Miftahutdinov, Andrey Sakhovskiy and Elena Tutubalina	51
<i>SMM4H Shared Task 2020 - A Hybrid Pipeline for Identifying Prescription Drug Abuse from Twitter: Machine Learning, Deep Learning, and Post-Processing</i> Isabel Metzger, Emir Y. Haskovic, Allison Black, Whitley M. Yi, Rajat S. Chandra, Mark T. Rutledge, William McMahon and Yindalon Aphinyanaphongs	57
<i>Automatic Detecting for Health-related Twitter Data with BioBERT</i> Yang Bai and Xiaobing Zhou	63
<i>Exploring Online Depression Forums via Text Mining: A Comparison of Reddit and a Curated Online Forum</i> Luis Moßburger, Felix Wende, Kay Brinkmann and Thomas Schmidt	70
<i>Towards Preemptive Detection of Depression and Anxiety in Twitter</i> David Owen, Jose Camacho-Collados and Luis Espinosa Anke	82
<i>Identifying Medication Abuse and Adverse Effects from Tweets: University of Michigan at #SMM4H 2020</i> V.G.Vinod Vydiswaran, Deahan Yu, Xinyan Zhao, Ermioni Carr, Jonathan Martindale, Jingcheng Xiao, Noha Ghannam, Matteo Althoen, Alexis Castellanos, Neel Patel and Daniel Vasquez	90
<i>How Far Can We Go with Just Out-of-the-box BERT Models?</i> Lucie Gattepaille	95

<i>FBK@SMM4H2020: RoBERTa for Detecting Medications on Twitter</i> Silvia Casola and Alberto Lavelli	101
<i>Autobots Ensemble: Identifying and Extracting Adverse Drug Reaction from Tweets Using Transformer Based Pipelines</i> Sougata Saha, Souvik Das, Prashi Khurana and Rohini Srihari	104
<i>Transformer Models for Drug Adverse Effects Detection from Tweets</i> Pavel Blinov and Manvel Avetisian	110
<i>Adverse Drug Reaction Detection in Twitter Using RoBERTa and Rules</i> Sedigh Khademi, Pari Delirhaghighi and Frada Burstein	113
<i>SpeechTrans@SMM4H'20: Impact of Preprocessing and N-grams on Automatic Classification of Tweets That Mention Medications</i> Mohamed Lichouri and Mourad Abbas	118
<i>Want to Identify, Extract and Normalize Adverse Drug Reactions in Tweets? Use RoBERTa</i> Katikapalli Subramanyam Kalyan and Sivanesan Sangeetha	121
<i>Detecting Tweets Reporting Birth Defect Pregnancy Outcome Using Two-View CNN RNN Based Architecture</i> Saichethan Reddy	125
<i>Identification of Medication Tweets Using Domain-specific Pre-trained Language Models</i> Yandrapati Prakash Babu and Rajagopal Eswari	128
<i>Medication Mention Detection in Tweets Using ELECTRA Transformers and Decision Trees</i> Lung-Hao Lee, Po-Han Chen, Hao-Chuan Kao, Ting-Chun Hung, Po-Lei Lee and Kuo-Kai Shyu	131
<i>LITL at SMM4H: An Old-school Feature-based Classifier for Identifying Adverse Effects in Tweets</i> Ludovic Tanguy, Lydia-Mai Ho-Dac, Cécile Fabre, Roxane Bois, Touati Mohamed Yacine Haddad, Claire Ibarboure, Marie Joyau, François Le moal, Jade Moilic, Laura Roudaut, Mathilde Simounet, Irena Stankovic and Mickaela Vandewaetere	134
<i>Sentence Classification with Imbalanced Data for Health Applications</i> Farhana Ferdousi Liza	138
<i>HITSZ-ICRC: A Report for SMM4H Shared Task 2020-Automatic Classification of Medications and Adverse Effect in Tweets</i> Xiaoyu Zhao, Ying Xiong and Buzhou Tang	146
<i>Automatic Classification of Tweets Mentioning a Medication Using Pre-trained Sentence Encoders</i> Laiba Mehnaz	150
<i>Approaching SMM4H 2020 with Ensembles of BERT Flavours</i> George-Andrei Dima, Andrei-Marius Avram and Dumitru-Clementin Cercel	153
<i>NLP@VCU: Identifying Adverse Effects in English Tweets for Unbalanced Data</i> Darshini Mahendran, Cora Lewis and Bridget McInnes	158
<i>Sentence Transformers and Bayesian Optimization for Adverse Drug Effect Detection from Twitter</i> Oguzhan Gencoglu	161

Sentence Contextual Encoder with BERT and BiLSTM for Automatic Classification with Imbalanced Medication Tweets

Olanrewaju Tahir Aduragba, Jialin Yu, Gautham Senthilnathan and Alexandra Crsitea 165

CLaC at SMM4H 2020: Birth Defect Mention Detection

Parsa Bagherzadeh and Sabine Bergler 168

Workshop Program

Saturday, December 12, 2020

14:00–14:10 *Introduction*
Graciela Gonzalez-Hernandez

14:10–14:50 **Oral Presentations Q&A Session 1**

COVID-19 Twitter Monitor: Aggregating and Visualizing COVID-19 Related Trends in Social Media

Joseph Cornelius, Tilia Ellendorff, Lenz Furrer and Fabio Rinaldi

Conversation-Aware Filtering of Online Patient Forum Messages
Anne Dirkson, Suzan Verberne and Wessel Kraaij

Annotating Patient Information Needs in Online Diabetes Forums

Julia Romberg, Jan Dyczmons, Sandra Olivia Borgmann, Jana Sommer, Markus Vomhof, Cecilia Brunoni, Ismael Bruck-Ramisch, Luis Enders, Andrea Icks and Stefan Conrad

Overview of the Fifth Social Media Mining for Health Applications (#SMM4H) Shared Tasks at COLING 2020

Ari Klein, Ilseyar Alimova, Ivan Flores, Arjun Magge, Zulfat Miftahutdinov, Anne-Lyse Minard, Karen O’Connor, Abeed Sarker, Elena Tutubalina, Davy Weisenbacher and Graciela Gonzalez-Hernandez

14:50–15:00 **Break**

15:00–15:40 *Invited Talk*
Fabio Rinaldi

15:40–15:50 **Break**

15:50–16:30 **Oral Presentations Q&A Session 2**

Ensemble BERT for Classifying Medication-mentioning Tweets

Huong Dang, Kahyun Lee, Sam Henry and Özlem Uzuner

ISLab System for SMM4H Shared Task 2020

Chen-Kai Wang, Hong-Jie Dai, You-Chen Zhang, Bo-Chun Xu, Bo-Hong Wang, You-Ning Xu, Po-Hao Chen and Chung-Hong Lee

BERT Implementation for Detecting Adverse Drug Effects Mentions in Russian

Andrey Gusev, Anna Kuznetsova, Anna Polyanskaya and Egor Yatsishin

KFU NLP Team at SMM4H 2020 Tasks: Cross-lingual Transfer Learning with Pre-trained Language Models for Drug Reactions

Zulfat Miftahutdinov, Andrey Sakhovskiy and Elena Tutubalina

Saturday, December 12, 2020 (continued)

16:30–17:30 **Poster Session**

17:30–17:40 **Break**

17:40–18:20 **Oral Presentations Q&A Session 3**

SMM4H Shared Task 2020 - A Hybrid Pipeline for Identifying Prescription Drug Abuse from Twitter: Machine Learning, Deep Learning, and Post-Processing

Isabel Metzger, Emir Y. Haskovic, Allison Black, Whitley M. Yi, Rajat S. Chandra, Mark T. Rutledge, William McMahon and Yindalon Aphinyanaphongs

Automatic Detecting for Health-related Twitter Data with BioBERT

Yang Bai and Xiaobing Zhou

Exploring Online Depression Forums via Text Mining: A Comparison of Reddit and a Curated Online Forum

Luis Moßburger, Felix Wende, Kay Brinkmann and Thomas Schmidt

Towards Preemptive Detection of Depression and Anxiety in Twitter

David Owen, Jose Camacho-Collados and Luis Espinosa Anke

18:20–18:30 *Conclusion*

Graciela Gonzalez-Hernandez