

NAACL-HLT 2021

Social Media Mining for Health (SMM4H)

Proceedings of the Sixth Workshop and Shared Tasks

June 10, 2021

©2021 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-954085-31-2

Preface

Welcome to the 6th Social Media Mining for Health (#SMM4H) Workshop & Shared Task 2021, co-located at the 2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Held online in its sixth iteration, #SMM4H 2021 continues to serve as a venue for bringing together data mining researchers interested in building solutions for challenges involved in utilizing social media data for health informatics. For #SMM4H 2021, we accepted 3 workshop papers and 29 shared task system description papers. Each submission was peer-reviewed by two to three reviewers.

The accepted workshop papers used social media data, mainly from Twitter, for topics, studies and applications surrounding COVID-19 and pharmacovigilance. Niu et. al. present a study summarizing the evaluation of Twitter sentiments towards non-pharmaceutical interventions for COVID-19 in Canada. Karisani et. al. propose a novel technique that uses both unlabeled and labeled tweets with drug mentions along multiple views to achieve a new state-of-the-art performance in extracting adverse drug effects. Finally, Miranda et. al. present a new annotated corpora in Spanish to identify occupational subgroups on Twitter to estimate risks associated with COVID-19. They also present a summary of the ProfNER shared task organized with the annotated data along the text classification and named entity recognition subtasks.

The #SMM4H 2021 shared tasks sought to advance the use of Twitter data (tweets) for pharmacovigilance, medication non-adherence, patient-centered outcomes, tracking cases and symptoms associated with COVID-19 and assessing risks for occupational groups. In addition to re-runs of adverse drug effects extraction tasks in English and Russian #SMM4H 2021 included new tasks for detecting medication non-adherence, adverse pregnancy outcomes, probable cases of COVID-19, symptoms associated with COVID-19, extracting occupations and professions from Spanish tweets for COVID-19 risk assessment and detecting self reports of breast cancer posts. The eight tasks required methods for binary classification, multi-class classification, and named entity recognition (NER). With 40 teams making prediction submissions, participation in the #SMM4H shared tasks continue to grow. Among the 29 shared task system description papers that were accepted, 9 teams were invited to present their system orally.

The organizing committee of #SMM4H 2021 would like to thank the program committee for reviewing the workshop papers and the additional reviewers of system description papers for providing constructive feedback and participating in peer-review. We are also grateful to the organizers of NAACL 2021 for facilitating the organization of the workshop and the Codalab team for providing the platform to organize shared tasks. We would also like to thank the annotators of the shared task datasets, and of course, everyone who submitted a paper or participated in the shared tasks. #SMM4H 2021 would not have been possible without the contributions and participation from all of them.

Arjun, Ari, Antonio, Mohammed Ali, Ilseyar, Zulfat, Eulàlia, Salvador, Ivan, Karen, Davy, Elena, Abeer, Juan, Martin and Graciela

Organizing and Program Committees

Organizing Committee:

Graciela Gonzalez-Hernandez, University of Pennsylvania, USA
Arjun Magge, University of Pennsylvania, USA
Ari Z. Klein, University of Pennsylvania, USA
Davy Weissenbacher, University of Pennsylvania, USA
Ivan Flores, University of Pennsylvania, USA
Karen O'Connor, University of Pennsylvania, USA
Martin Krallinger, Barcelona Supercomputing Center, Spain
Antonio Miranda-Escalada, Barcelona Supercomputing Center, Spain
Eulàlia Farré-Maduell, Barcelona Supercomputing Center, Spain
Salvador Lima López, Barcelona Supercomputing Center, Spain
Juan M. Banda, Georgia State University, USA
Abeed Sarker, Emory University, USA
Mohammed Ali Al-garadi, Emory University, USA
Elena Tutubalina, Kazan Federal University, Russia
Ilseyar Alimova, Kazan Federal University, Russia
Zulfat Miftahutdinov, Kazan Federal University, Russia

Program Committee:

Olivier Bodenreider, US National Library of Medicine, USA
Pierre Zweigenbaum, French National Center for Scientific Research, France
Kirk Roberts, University of Texas Health Science Center at Houston, USA
Rajesh Piryani, South Asian University, India
Yutaka Sasaki, Toyota Technological Institute, Japan
Nicolas Turenne, French National Institute for Agricultural Research, France

Table of Contents

<i>Statistically Evaluating Social Media Sentiment Trends towards COVID-19 Non-Pharmaceutical Interventions with Event Studies</i>	
Jingcheng Niu, Erin Rees, Victoria Ng and Gerald Penn	1
<i>View Distillation with Unlabeled Data for Extracting Adverse Drug Effects from User-Generated Data</i>	
Payam Karisani, Jinho D. Choi and Li Xiong	7
<i>The ProfNER shared task on automatic recognition of occupation mentions in social media: systems, evaluation, guidelines, embeddings and corpora</i>	
Antonio Miranda-Escalada, Eulàlia Farré-Maduell, Salvador Lima López, Luis Gascó, Vicent Briva-Iglesias, Marvin Agüero-Torales and Martin Krallinger	25
<i>Overview of the Sixth Social Media Mining for Health Applications (#SMM4H) Shared Tasks at NAACL 2021</i>	
Arjun Magge, Ari Klein, Antonio Miranda-Escalada, Mohammed Ali Al-Garadi, Ilseyar Alimova, Zulfat Miftahutdinov, Eulalia Farre, Salvador Lima López, Ivan Flores, Karen O'Connor, Davy Weissenbacher, Elena Tutubalina, Abeed Sarker, Juan Banda, Martin Krallinger and Graciela Gonzalez-Hernandez	13
<i>BERT based Transformers lead the way in Extraction of Health Information from Social Media</i>	
Sidharth Ramesh, Abhiraj Tiwari, Parthivi Choubey, Saisha Kashyap, Sahil Khose, Kumud Lakara, Nishesh Singh and Ujjwal Verma	33
<i>KFU NLP Team at SMM4H 2021 Tasks: Cross-lingual and Cross-modal BERT-based Models for Adverse Drug Effects</i>	
Andrey Sakhovskiy, Zulfat Miftahutdinov and Elena Tutubalina	39
<i>Transformer-based Multi-Task Learning for Adverse Effect Mention Analysis in Tweets</i>	
George-Andrei Dima, Dumitru-Clementin Cercel and Mihai Dascalu	44
<i>Pre-trained Transformer-based Classification and Span Detection Models for Social Media Health Applications</i>	
Yuting Guo, Yao Ge, Mohammed Ali Al-Garadi and Abeed Sarker	52
<i>BERT Goes Brrr: A Venture Towards the Lesser Error in Classifying Medical Self-Reporters on Twitter</i>	
Alham Fikri Aji, Made Nindyatama Nityasya, Haryo Akbarianto Wibowo, Radityo Eko Prasajo and Tirana Fatyanosa	58
<i>UACH-INAOE at SMM4H: a BERT based approach for classification of COVID-19 Twitter posts</i>	
Alberto Valdes, Jesus Lopez and Manuel Montes	65
<i>System description for ProfNER - SMMH: Optimized finetuning of a pretrained transformer and word vectors</i>	
David Carreto Fidalgo, Daniel Vila-Suero, Francisco Aranda Montes and Ignacio Talavera Cepeda	69
<i>Word Embeddings, Cosine Similarity and Deep Learning for Identification of Professions & Occupations in Health-related Social Media</i>	
Sergio Santamaría Carrasco and Roberto Cuervo Rosillo	74

<i>Classification, Extraction, and Normalization : CASIA_Unisound Team at the Social Media Mining for Health 2021 Shared Tasks</i>	
Tong Zhou, Zhucong Li, Zhen Gan, Baoli Zhang, Yubo Chen, Kun Niu, Jing Wan, Kang Liu, Jun Zhao, Yafei Shi, Weifeng Chong and Shengping Liu	77
<i>Neural Text Classification and Stacked Heterogeneous Embeddings for Named Entity Recognition in SMM4H 2021</i>	
Usama Yaseen and Stefan Langer	83
<i>BERT based Adverse Drug Effect Tweet Classification</i>	
Tanay Kayastha, Pranjal Gupta and Pushpak Bhattacharyya	88
<i>A Joint Training Approach to Tweet Classification and Adverse Effect Extraction and Normalization for SMM4H 2021</i>	
Mohab Elkaref and Lamiece Hassan	91
<i>Text Augmentation Techniques in Drug Adverse Effect Detection Task</i>	
Pavel Blinov	95
<i>Classification of Tweets Self-reporting Adverse Pregnancy Outcomes and Potential COVID-19 Cases Using RoBERTa Transformers</i>	
Lung-Hao Lee, Man-Chen Hung, Chien-Huan Lu, Chang-Hao Chen, Po-Lei Lee and Kuo-Kai Shyu	98
<i>NLP@NISER: Classification of COVID19 tweets containing symptoms</i>	
Deepak Kumar, Nalin Kumar and Subhankar Mishra	102
<i>Identification of profession & occupation in Health-related Social Media using tweets in Spanish</i>	
Victoria Pachón, Jacinto Mata Vázquez and Juan Luís Domínguez Olmedo	105
<i>Lasige-BioTM at ProfNER: BiLSTM-CRF and contextual Spanish embeddings for Named Entity Recognition and Tweet Binary Classification</i>	
Pedro Ruas, Vitor Andrade and Francisco Couto	108
<i>Adversities are all you need: Classification of self-reported breast cancer posts on Twitter using Adversarial Fine-tuning</i>	
Adarsh Kumar, OJASV Kamal and Susmita Mazumdar	112
<i>UoB at ProfNER 2021: Data Augmentation for Classification Using Machine Translation</i>	
Frances Adriana Laureano De Leon, Harish Tayyar Madabushi and Mark Lee	115
<i>Transformer Models for Classification on Health-Related Imbalanced Twitter Datasets</i>	
Varad Pimpalkhute, Prajwal Nakhate and Tausif Diwan	118
<i>OCHADAI at SMM4H-2021 Task 5: Classifying self-reporting tweets on potential cases of COVID-19 by ensembling pre-trained language models</i>	
Ying Luo, Lis Pereira and Kobayashi Ichiro	123
<i>PAII-NLP at SMM4H 2021: Joint Extraction and Normalization of Adverse Drug Effect Mentions in Tweets</i>	
Zongcheng Ji, Tian Xia and Mei Han	126
<i>Assessing multiple word embeddings for named entity recognition of professions and occupations in health-related social media</i>	
Vasile Pais and Maria Mitrofan	128

<i>Fine-tuning Transformers for Identifying Self-Reporting Potential Cases and Symptoms of COVID-19 in Tweets</i>	
Max Fleming, Priyanka Dondeti, Caitlin Dreisbach and Adam Poliak	131
<i>Classification of COVID19 tweets using Machine Learning Approaches</i>	
Anupam Mondal, Sainik Mahata, Monalisa Dey and Dipankar Das	135
<i>Fine-tuning BERT to classify COVID19 tweets containing symptoms</i>	
Rajarshi Roychoudhury and Sudip Naskar	138
<i>Identifying professions & occupations in Health-related Social Media using Natural Language Processing</i>	
Alberto Mesa Murgado, Ana Parras Portillo, Pilar López Úbeda, Maite Martin and Alfonso Ureña-López	141
<i>Approaching SMM4H with auto-regressive language models and back-translation</i>	
Joseph Cornelius, Tilia Ellendorff and Fabio Rinaldi	146
<i>ULD-NUIG at Social Media Mining for Health Applications (#SMM4H) Shared Task 2021</i>	
Atul Kr. Ojha, Priya Rani, Koustava Goswami, Bharathi Raja Chakravarthi and John P. McCrae	149

Conference Program

June 10th 2021

09:00–09:15 *Opening Remarks and Introduction*
Graciela Gonzalez-Hernandez

09:15–10:15 Oral Presentations Q&A Session 1

Statistically Evaluating Social Media Sentiment Trends towards COVID-19 Non-Pharmaceutical Interventions with Event Studies

Jingcheng Niu, Erin Rees, Victoria Ng and Gerald Penn

View Distillation with Unlabeled Data for Extracting Adverse Drug Effects from User-Generated Data

Payam Karisani, Jinho D. Choi and Li Xiong

Overview of the Sixth Social Media Mining for Health Applications (#SMM4H) Shared Tasks at NAACL 2021

Arjun Magge, Ari Klein, Antonio Miranda-Escalada, Mohammed Ali Al-Garadi, Ilseyar Alimova, Zulfat Miftahutdinov, Eulàlia Farré-Maduell, Salvador Lima López, Ivan Flores, Karen O'Connor, Davy Weissenbacher, Elena Tutubalina, Abeed Sarker, Juan Banda, Martin Krallinger and Graciela Gonzalez-Hernandez

The ProfNER shared task on automatic recognition of occupation mentions in social media: systems, evaluation, guidelines, embeddings and corpora

Antonio Miranda-Escalada, Eulàlia Farré-Maduell, Salvador Lima López, Luis Gascó, Vicent Briva-Iglesias, Marvin Agüero-Torales and Martin Krallinger

10:15–10:30 **Break**

10:30–11:10 *Invited Talk*
Mark Dredze

11:10–11:25 **Break**

June 10th 2021 (continued)

11:30–12:30 Oral Presentations Q&A Session 2

BERT based Transformers lead the way in Extraction of Health Information from Social Media

Sidharth Ramesh, Abhiraj Tiwari, Parthivi Choubey, Saisha Kashyap, Sahil Khose, Kumud Lakara, Nishesh Singh and Ujjwal Verma

KFU NLP Team at SMM4H 2021 Tasks: Cross-lingual and Cross-modal BERT-based Models for Adverse Drug Effects

Andrey Sakhovskiy, Zulfat Miftahutdinov and Elena Tutubalina

Transformer-based Multi-Task Learning for Adverse Effect Mention Analysis in Tweets

George-Andrei Dima, Dumitru-Clementin Cercel and Mihai Dascalu

Pre-trained Transformer-based Classification and Span Detection Models for Social Media Health Applications

Yuting Guo, Yao Ge, Mohammed Ali Al-Garadi and Abeed Sarker

12:30–13:15 Poster Session

13:15–13:30 Break

13:30–14:45 Oral Presentations Q&A Session 3

BERT Goes Brrr: A Venture Towards the Lesser Error in Classifying Medical Self-Reporters on Twitter

Alham Fikri Aji, Made Nindyatama Nityasya, Haryo Akbarianto Wibowo, Radityo Eko Prasajo and Tirana Fatyanosa

UACH-INAOE at SMM4H: a BERT based approach for classification of COVID-19 Twitter posts

Alberto Valdes, Jesus Lopez and Manuel Montes

System description for ProfNER - SMMH: Optimized finetuning of a pretrained transformer and word vectors

David Carreto Fidalgo, Daniel Vila-Suero, Francisco Aranda Montes and Ignacio Talavera Cepeda

Word Embeddings, Cosine Similarity and Deep Learning for Identification of Professions & Occupations in Health-related Social Media

Sergio Santamaría Carrasco and Roberto Cuervo Rosillo

June 10th 2021 (continued)

Classification, Extraction, and Normalization : CASIA_Unisound Team at the Social Media Mining for Health 2021 Shared Tasks

Tong Zhou, Zhucong Li, Zhen Gan, Baoli Zhang, Yubo Chen, Kun Niu, Jing Wan, Kang Liu, Jun Zhao, Yafei Shi, Weifeng Chong and Shengping Liu

14:45–15:00 *Conclusion and Closing Remarks*
Graciela Gonzalez-Hernandez

