

# **6th Social Media Mining for Health Workshop and Shared Tasks (SMM4H 2021)**

Online  
10 June 2021

ISBN: 978-1-7138-3029-0

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2021) by the Association for Computational Linguistics  
All rights reserved.

Printed with permission by Curran Associates, Inc. (2021)

For permission requests, please contact the Association for Computational Linguistics  
at the address below.

Association for Computational Linguistics  
209 N. Eighth Street  
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006  
Fax: 1-570-476-0860

[acl@aclweb.org](mailto:acl@aclweb.org)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## 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 .....	13
<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 .....	21
<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 Prasojo 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 &amp; 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 &amp; 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 &amp; 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