

COLING 2022  
Volume 29 (2022), No. 18

**Proceedings of The Seventh Workshop on Social Media  
Mining for Health Applications, Workshop & Shared Task  
(#SMM4H 2022)**

**The 29th International Conference on Computational  
Linguistics**

October 12 - 17, 2022  
Gyeongju, Republic of Korea

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

ISSN 2951-2093

## Preface

Welcome to the 7th Social Media Mining for Health Applications (#SMM4H) Workshop & Shared Task 2022, co-located at the 29th International Conference on Computational Linguistics. Held as a hybrid event in its seventh iteration, #SMM4H 2022 continues to serve as a unique venue for bringing together data mining researchers interested in building and sharing solutions for utilizing social media data for health informatics. For #SMM4H 2022, we accepted 6 workshop papers and 47 shared task system description papers. Each submission was peer-reviewed by two or three reviewers.

The accepted papers proposed advanced models to detect or extract health-related information in posts written in various languages. Pais et al. report the performance of baseline transformers on a new corpus of Romanian micro-blogging posts annotated with 9 entity classes. Their corpus is made available to the community. Zanwar et al. detect 6 mental health conditions on Reddit posts with an interpretable neural network by combining a feature-based model with a transformer model. Chan et al. describe the collection and annotation process to create a corpus of Dutch Facebook comments. These posts comment on news articles about COVID-19 vaccination where Facebook users shared the knowledge they acquired through their personal experiences. Adhikari et al. detailed their GUI to improve, with incremental learning, their classifier of 8 topics related to COVID-19 in Nepali tweets. Finally, Davydova & Tutubalina and Gasco Sánchez et al. expand the description and the analysis of the results of the #SMM4H shared tasks 2 and 10.

The #SMM4H 2022 shared tasks sought to advance the use of user-generated social media data for pharmacovigilance, epidemiology, patient-centered outcomes, and tracking beliefs and impacts of COVID-19. #SMM4H 2022 included re-runs of three tasks about adverse drug events, changes in medication treatments, and COVID-19 symptoms. In addition, #SMM4H 2022 included seven new tasks on detecting stances toward COVID-19 health mandates, COVID-19 vaccination status, the age of social media users, intimate partner violence, chronic stress, and diseases. The ten tasks required methods for multi-class classification, and named entity recognition and normalization. With 117 teams that registered from 28 countries and 54 teams that participated, the interest in the #SMM4H shared tasks continues to grow. Among the 47 system description papers that were accepted, 10 teams were invited for an oral presentation.

The organizing committee of #SMM4H 2022 would like to thank the program committee, the additional reviewers of system description papers, the organizers of COLING 2022 (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 2022 would not have been possible without them.

Graciela, Davy, Arjun, Ari, Ivan, Karen, Raul, Lucia, Juan, Abeed, Yuting, Yao, Elena, Luis, Darryl, and Martin.



## Organizing and Program Committees

### Organizing Committee:

Graciela Gonzalez-Hernandez, Cedars-Sinai Medical Center, USA  
Davy Weissenbacher, Cedars-Sinai Medical Center, USA  
Arjun Magge, University of Pennsylvania, USA  
Ari Z. Klein, University of Pennsylvania, USA  
Ivan Flores, Cedars-Sinai Medical Center, USA  
Karen O'Connor, University of Pennsylvania, USA  
Raul Rodriguez-Esteban, Roche Pharmaceuticals, Switzerland  
Lucia Schmidt, Roche Pharmaceuticals, Switzerland  
Juan M. Banda, Georgia State University, USA  
Abeed Sarker, Emory University, USA  
Yuting Guo, Emory University, USA  
Yao Ge, Emory University, USA  
Elena Tutubalina, Insilico Medicine, Hong Kong  
Luis Gasco, Barcelona Supercomputing Center, Spain  
Darryl Estrada, Barcelona Supercomputing Center, Spain  
Martin Krallinger, Barcelona Supercomputing Center, Spain

### Program Committee:

Cecilia Arighi, University of Delaware, USA  
Natalia Grabar, French National Center for Scientific Research, France  
Thierry Hamon, Paris-Nord University, France  
Antonio Jimeno Yepes, Royal Melbourne Institute of Technology, Australia  
Jin-Dong Kim, Database Center for Life Science, Japan  
Corrado Lanera, University of Padova, Italy  
Robert Leaman, US National Library of Medicine, USA  
Kirk Roberts, University of Texas Health Science Center at Houston, USA  
Yutaka Sasaki, Toyota Technological Institute, Japan  
Pierre Zweigenbaum, French National Center for Scientific Research, France



# Conference Program

**#SMM4H'22 Monday, October 17, 2022**

9:00–9:15 *Introduction*  
Graciela Gonzalez-Hernandez

09:15–10:15 **Oral Presentations Q&A Session 1**  
*Romanian micro-blogging named entity recognition including health-related entities*  
Vasile Pais, Verginica Barbu Mititelu, Elena Irimia, Maria Mitrofan, Carol Luca Gasan and Roxana Micu

*The Best of Both Worlds: Combining Engineered Features with Transformers for Improved Mental Health Prediction from Reddit Posts*  
Sourabh Satish Zanwar, Daniel Wiechmann, Yu Qiao and Elma Kerz

*Overview of the Seventh Social Media Mining for Health Applications (#SMM4H) Shared Tasks at COLING 2022*  
Davy Weissenbacher, Juan Banda, Vera Davydova, Darryl Johan Estrada Zavala, Luis Gasco Sánchez, Yao Y. Ge, Yuting Guo, Ari Klein, Martin Krallinger, Mathias Leddin, Arjun Magge, Raul Rodriguez-Esteban, Abeed Sarker, Lucia L. Schmidt, Elena Tutubalina and Graciela Gonzalez-Hernandez

*PingAnTech at SMM4H task1: Multiple pre-trained model approaches for Adverse Drug Reactions*  
Xi Liu, Han Zhou and Chang Su

10:15–10:30 *Break*

10:30–11:30 **Oral Presentations Q&A Session 2**  
*COVID-19-related Nepali Tweets Classification in a Low Resource Setting*  
Rabin Adhikari, Safal Thapaliya, Nirajan Basnet, Samip Poudel, Aman Shakya and Bishesh Khanal

*Leveraging Social Media as a Source for Clinical Guidelines: A Demarcation of Experiential Knowledge*  
Jia-Zhen Michelle Chan, Florian Kunneman, Roser Morante, Lea Lösch and Teun Zuiderent-Jerak

*Zhegu@SMM4H-2022: The Pre-training Tweet & Claim Matching Makes Your Prediction Better*  
Pan He, Chen YuZe and Yanru Zhang

*CSECU-DSG@SMM4H'22: Transformer based Unified Approach for Classification of Changes in Medication Treatments in Tweets and WebMD Reviews*  
Afrin Sultana, Nihad Karim Chowdhury and Abu Nowshed Chy

**#SMM4H'22 Monday, October 17, 2022 (continued)**

11:30–12:15 **Poster Session**

12:15–12:30 *Break*

12:30–13:30 **Oral Presentations Q&A Session 3**

*Yet@SMM4H'22: Improved BERT-based classification models with Rdrop and PolyLoss*

Yan Zhuang and Yanru Zhang

*AIR-JPMC@SMM4H'22: Identifying Self-Reported Spanish COVID-19 Symptom Tweets Through Multiple-Model Ensembling*

Adrian Garcia Hernandez, Leung Wai Liu, Akshat Gupta, vineeth ravi, Saheed O. Obitayo, Xiaomo Liu and Sameena Shah

*AILAB-Udine@SMM4H'22: Limits of Transformers and BERT Ensembles*

Beatrice Portelli, Simone Scabro, Emmanuele Chersoni, Enrico Santus and Giuseppe Serra

*AIR-JPMC@SMM4H'22: Classifying Self-Reported Intimate Partner Violence in Tweets with Multiple BERT-based Models*

Alec Louis Clemente Candidato, Akshat Gupta, Xiaomo Liu and Sameena Shah

13:30–13:45 *Break*

13:45–14:30 **Oral Presentations Q&A Session 4**

*zydhjh4593@SMM4H'22: A Generic Pre-trained BERT-based Framework for Social Media Health Text Classification*

Chenghao Huang, Xiaolu Chen, Yuxi Chen, Yutong Wu, Weimin Yuan, Yan Wang and Yanru Zhang

*Fraunhofer FKIE @ SMM4H 2022: System Description for Shared Tasks 2, 4 and 9*

Daniel Claeser and Samantha Kent

*CASIA@SMM4H'22: A Uniform Health Information Mining System for Multilingual Social Media Texts*

Jia Fu, Sirui Li, hui ming yuan, Zhucong Li, zhen gan, Yubo Chen, kang liu, Jun Zhao and Shengping Liu

14:30–14:45 *Break*

14:45–15:25 **Keynote**

Raul Rodriguez-Esteban

15:25–15:40 *Conclusion and Closing Remarks*

Graciela Gonzalez-Hernandez



## Table of Contents

<i>RACAI@SMM4H'22: Tweets Disease Mention Detection Using a Neural Lateral Inhibitory Mechanism</i> Andrei-Marius Avram, Vasile Pais and Maria Mitrofan .....	1
<i>PingAnTech at SMM4H task1: Multiple pre-trained model approaches for Adverse Drug Reactions</i> Xi Liu, Han Zhou and Chang Su .....	4
<i>dezzai@SMM4H'22: Tasks 5 &amp; 10 - Hybrid models everywhere</i> Miguel Ortega-Martín, Alfonso Ardoiz, Oscar Garcia, Jorge Álvarez and Adrián Alonso .....	7
<i>zydhjh4593@SMM4H'22: A Generic Pre-trained BERT-based Framework for Social Media Health Text Classification</i> Chenghao Huang, Xiaolu Chen, Yuxi Chen, Yutong Wu, Weimin Yuan, Yan Wang and Yanru Zhang 11	
<i>MANTIS at SMM4H'2022: Pre-Trained Language Models Meet a Suite of Psycholinguistic Features for the Detection of Self-Reported Chronic Stress</i> Sourabh Zanwar, Daniel Wiechmann, Yu Qiao and Elma Kerz .....	16
<i>NLP-CIC-WFU at SocialDisNER: Disease Mention Extraction in Spanish Tweets Using Transfer Learning and Search by Propagation</i> Antonio Tamayo, Alexander Gelbukh and Diego Burgos .....	19
<i>yiriyu@SMM4H'22: Stance and Premise Classification in Domain Specific Tweets with Dual-View Attention Neural Networks</i> Huabin Yang, Zhongjian Zhang and Yanru Zhang .....	23
<i>SINAI@SMM4H'22: Transformers for biomedical social media text mining in Spanish</i> Mariia Chizhikova, Pilar López-Úbeda, Manuel C. Díaz-Galiano, L. Alfonso Ureña-López and M. Teresa Martín-Valdivia .....	27
<i>BOUN-TABI@SMM4H'22: Text-to-Text Adverse Drug Event Extraction with Data Balancing and Prompting</i> Gökçe Uludoğan and Zeynep Yirmibeşoğlu .....	31
<i>uestcc@SMM4H'22: RoBERTa based Adverse Drug Events Classification on Tweets</i> Chunchen Wei, Ran Bi and Yanru Zhang .....	35
<i>Zhegu@SMM4H-2022: The Pre-training Tweet &amp; Claim Matching Makes Your Prediction Better</i> Pan He, Chen YuZe and Yanru Zhang .....	38
<i>MaNLP@SMM4H'22: BERT for Classification of Twitter Posts</i> Keshav Kapur, Rajitha Harikrishnan and Sanjay Singh .....	42
<i>John_Snow_Labs@SMM4H'22: Social Media Mining for Health (#SMM4H) with Spark NLP</i> Veysel Kocaman, Cabir Celik, Damla Gurbaz, Gursev Pirge, Bunyamin Polat, Halil Saglamlar, Meryem Vildan Sarikaya, Gokhan Turer and David Talby .....	44
<i>READ-BioMed@SocialDisNER: Adaptation of an Annotation System to Spanish Tweets</i> Antonio Jimeno Yepes and Karin Verspoor .....	48

<i>PLN CMM at SocialDisNER: Improving Detection of Disease Mentions in Tweets by Using Document-Level Features</i>	
Matias Rojas, Jose Barros, Kinan Martin, Mauricio Araneda-Hernandez and Jocelyn Dunstan . .	52
<i>CLaCLab at SocialDisNER: Using Medical Gazetteers for Named-Entity Recognition of Disease Mentions in Spanish Tweets</i>	
Harsh Verma, Parsa Bagherzadeh and Sabine Bergler . . . . .	55
<i>CIC NLP at SMM4H 2022: a BERT-based approach for classification of social media forum posts</i>	
Atnafu Lambebo Tonja, Olumide Ebenezer Ojo, Mohammed Arif Khan, Abdul Gafar Manuel Meque, Olga Kolesnikova, Grigori Sidorov and Alexander Gelbukh . . . . .	58
<i>NCUEE-NLP@SMM4H'22: Classification of Self-reported Chronic Stress on Twitter Using Ensemble Pre-trained Transformer Models</i>	
Tzu-Mi Lin, Chao-Yi Chen, Yu-Wen Tzeng and Lung-Hao Lee . . . . .	62
<i>BioInfo@UAVR@SMM4H'22: Classification and Extraction of Adverse Event mentions in Tweets using Transformer Models</i>	
Edgar Morais, José Luis Oliveira, Alina Trifan and Olga Fajarda . . . . .	65
<i>FRE at SocialDisNER: Joint Learning of Language Models for Named Entity Recognition</i>	
Kendrick Cetina and Nuria García-Santa . . . . .	68
<i>ITAINNOVA at SocialDisNER: A Transformers cocktail for disease identification in social media in Spanish</i>	
Rosa Montañés-Salas, Irene López-Bosque, Luis García-Garcés and Rafael del-Hoyo-Alonso . .	71
<i>mattica@SMM4H'22: Leveraging sentiment for stance &amp; premise joint learning</i>	
Oscar Lithgow-Serrano, Joseph Cornelius, Fabio Rinaldi and Ljiljana Dolamic . . . . .	75
<i>KU_ED at SocialDisNER: Extracting Disease Mentions in Tweets Written in Spanish</i>	
Antoine Lain, Wonjin Yoon, Hyunjae Kim, Jaewoo Kang and Ian Simpson . . . . .	78
<i>CHAAI@SMM4H'22: RoBERTa, GPT-2 and Sampling - An interesting concoction</i>	
Christopher Palmer, Sedigheh Khademi Habibabadi, Muhammad Javed, Gerardo Luis Dimaguila and Jim Buttery . . . . .	81
<i>IAI @ SocialDisNER : Catch me if you can! Capturing complex disease mentions in tweets</i>	
Aman Sinha, Cristina Garcia Holgado, Marianne Clausel and Matthieu Constant . . . . .	85
<i>UCCNLP@SMM4H'22: Label distribution aware long-tailed learning with post-hoc posterior calibration applied to text classification</i>	
Paul Trust, Provia Kadusabe, Ahmed Zahran, Rosane Minghim and Kizito Omala . . . . .	90
<i>HaleLab_NITK@SMM4H'22: Adaptive Learning Model for Effective Detection, Extraction and Normalization of Adverse Drug Events from Social Media Data</i>	
Reshma Unnikrishnan, Sowmya Kamath S and Ananthanarayana V. S. . . . .	95
<i>Yet@SMM4H'22: Improved BERT-based classification models with Rdrop and PolyLoss</i>	
Yan Zhuang and Yanru Zhang . . . . .	98
<i>Fraunhofer FKIE @ SMM4H 2022: System Description for Shared Tasks 2, 4 and 9</i>	
Daniel Claeser and Samantha Kent . . . . .	103

<i>Transformer-based classification of premise in tweets related to COVID-19</i> Vadim Porvatov and Natalia Semenova .....	108
<i>Fraunhofer SIT@SMM4H'22: Learning to Predict Stances and Premises in Tweets related to COVID-19 Health Orders Using Generative Models</i> Raphael Frick and Martin Steinebach .....	111
<i>UB Health Miners@SMM4H'22: Exploring Pre-processing Techniques To Classify Tweets Using Transformer Based Pipelines.</i> Roshan Khatri, Sougata Saha, Souvik Das and Rohini Srihari .....	114
<i>CSECU-DSG@SMM4H'22: Transformer based Unified Approach for Classification of Changes in Medication Treatments in Tweets and WebMD Reviews</i> Afrin Sultana, Nihad Karim Chowdhury and Abu Nowshed Chy.....	118
<i>Innovators @ SMM4H'22: An Ensembles Approach for self-reporting of COVID-19 Vaccination Status Tweets</i> Mohammad Zohair, Nidhir Bhavsar, Aakash Bhatnagar and Muskaan Singh .....	123
<i>Innovators@SMM4H'22: An Ensembles Approach for Stance and Premise Classification of COVID-19 Health Mandates Tweets</i> Vatsal Savaliya, Aakash Bhatnagar, Nidhir Bhavsar and Muskaan Singh.....	126
<i>AILAB-Udine@SMM4H'22: Limits of Transformers and BERT Ensembles</i> Beatrice Portelli, Simone Scaboro, Emmanuele Chersoni, Enrico Santus and Giuseppe Serra ..	130
<i>AIR-JPMC@SMM4H'22: Classifying Self-Reported Intimate Partner Violence in Tweets with Multiple BERT-based Models</i> Alec Louis Candidato, Akshat Gupta, Xiaomo Liu and Sameena Shah .....	135
<i>ARGUABLY@SMM4H'22: Classification of Health Related Tweets using Ensemble, Zero-Shot and Fine-Tuned Language Model</i> Prabsimran Kaur, Guneet Kohli and Jatin Bedi.....	138
<i>CASIA@SMM4H'22: A Uniform Health Information Mining System for Multilingual Social Media Texts</i> Jia Fu, Sirui Li, hui ming yuan, Zhucong Li, zhen gan, Yubo Chen, kang liu, Jun Zhao and Shengping Liu .....	143
<i>Edinburgh_UCL_Health@SMM4H'22: From Glove to Flair for handling imbalanced healthcare corpora related to Adverse Drug Events, Change in medication and self-reporting vaccination</i> Imane Guellil, Jinge Wu, Honghan Wu, Tony Sun and Beatrice Alex.....	148
<i>KUL@SMM4H'22: Template Augmented Adaptive Pre-training for Tweet Classification</i> Sumam Francis and Marie-Francine Moens .....	153
<i>Enolp musk@SMM4H'22 : Leveraging Pre-trained Language Models for Stance And Premise Classification</i> Millon Das, Archit Mangrulkar, Ishan Manchanda, Manav Kapadnis and Sohan Patnaik .....	156
<i>AIR-JPMC@SMM4H'22: Identifying Self-Reported Spanish COVID-19 Symptom Tweets Through Multiple-Model Ensembling</i> Adrian Garcia Hernandez, Leung Wai Liu, Akshat Gupta, vineeth ravi, Saheed O. Obitayo, Xiaomo Liu and Sameena Shah.....	160

<i>AIR-JPMC@SMM4H'22: BERT + Ensembling = Too Cool: Using Multiple BERT Models Together for Various COVID-19 Tweet Identification Tasks</i>	
Leung Wai Liu, Akshat Gupta, Saheed Obitayo, Xiaomo Liu and Sameena Shah . . . . .	163
<i>CAISA@SMM4H'22: Robust Cross-Lingual Detection of Disease Mentions on Social Media with Adversarial Methods</i>	
Akbar Karimi and Lucie Flek . . . . .	168
<i>OFU@SMM4H'22: Mining Advent Drug Events Using Pretrained Language Models</i>	
Omar Adjali, Fréjus A. A. Laleye and Umang Aggarwal . . . . .	171
<i>CompLx@SMM4H'22: In-domain pretrained language models for detection of adverse drug reaction mentions in English tweets</i>	
Orest Xherija and Hojoon Choi . . . . .	176
<i>The SocialDisNER shared task on detection of disease mentions in health-relevant content from social media: methods, evaluation, guidelines and corpora</i>	
Luis Gasco Sánchez, Darryl Estrada Zavala, Eulàlia Farré-Maduell, Salvador Lima-López, Antonio Miranda-Escalada and Martin Krallinger . . . . .	182
<i>Romanian micro-blogging named entity recognition including health-related entities</i>	
Vasile Pais, Verginica Barbu Mititelu, Elena Irimia, Maria Mitrofan, Carol Luca Gasan and Roxana Micu . . . . .	190
<i>The Best of Both Worlds: Combining Engineered Features with Transformers for Improved Mental Health Prediction from Reddit Posts</i>	
Sourabh Zanwar, Daniel Wiechmann, Yu Qiao and Elma Kerz . . . . .	197
<i>Leveraging Social Media as a Source for Clinical Guidelines: A Demarcation of Experiential Knowledge</i>	
Jia-Zhen Michelle Chan, Florian Kunneman, Roser Morante, Lea Lösch and Teun Zuiderent-Jerak	203
<i>COVID-19-related Nepali Tweets Classification in a Low Resource Setting</i>	
Rabin Adhikari, Safal Thapaliya, Nirajan Basnet, Samip Poudel, Aman Shakya and Bishesh Khanal	209
<i>SMM4H 2022 Task 2: Dataset for stance and premise detection in tweets about health mandates related to COVID-19</i>	
Vera Davydova and Elena Tutubalina . . . . .	216
<i>Overview of the Seventh Social Media Mining for Health Applications (#SMM4H) Shared Tasks at COLING 2022</i>	
Davy Weissenbacher, Juan Banda, Vera Davydova, Darryl Estrada Zavala, Luis Gasco Sánchez, Yao Ge, Yuting Guo, Ari Klein, Martin Krallinger, Mathias Leddin, Arjun Magge, Raul Rodriguez-Esteban, Abeer Sarker, Lucia Schmidt, Elena Tutubalina and Graciela Gonzalez-Hernandez . . . . .	221