Generalizable Natural Language Processing Framework for Migraine Reporting from Social Media

Yuting Guo[1], **Swati Rajwal**[1], Sahithi Lakamana[1], Chia-Chun Chiang[2], Paul C. Menell[2], Adnan H. Shahid[2], Yi-Chieh Chen[3], Nikita Chhabra[4], Wan-Ju Chao[5], Chieh-Ju Chao[6], Todd J. Schwedt[4], Imon Banerjee[7,8] and Abeed Sarker[1]

[1]Biomedical Informatics, Emory University, GA, [2]Neurology dept., Mayo Clinic, MN, [3]Pharmacy Services Dept., Mayo Clinic, MN, [4]Neurology Dept., Mayo Clinic, AZ, [5]Psychology Dept., University of North Texas, TX, [6]Cardiology Dept., Mayo Clinic, MN, [7]Radiology Dept., Mayo Clinic, AZ, [8]CS Dept., Arizona State University, AZ

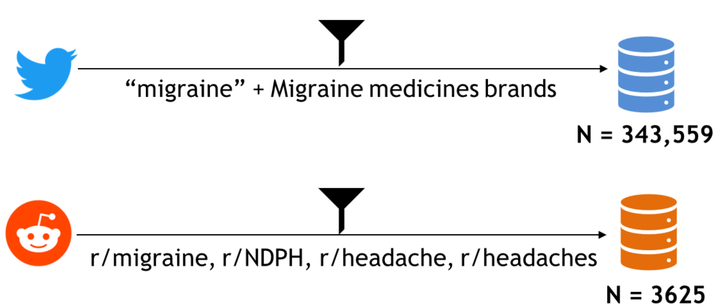
Abstract

- Our objectives:
  - ➢ Verify existence of self-reported migraine chatter on social media
  - ➢ Develop supervised text classifier for detecting self-reported migraine post
  - ➢ Assess the utility of social media for studying cohort-specific challenges.
- Expert-annotated **5750** Tweets & **302** Reddit posts
- Best system F_1 score **0.90** (Twitter), **0.93** (Reddit)
- Analysis show sentiment trends associated with migraine medications

Background

- **Patient-generated social media data** captures daily **habits/interests**[1]
- Social media with NLP improve patient-centered outcomes in cohort studies
- Example: breast cancer, substance use[2,3]
- Studies[4] investigated migraine using social media.
- Unclear if such methods are portable to other social media platforms

Dataset



Methodology

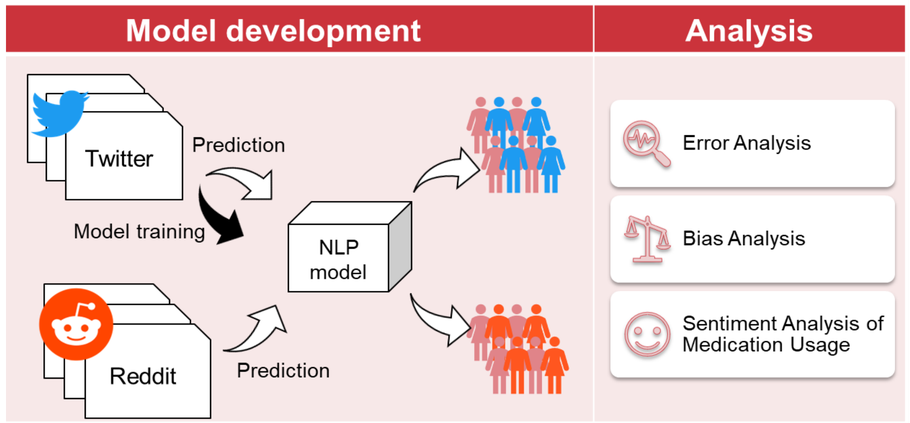


Figure 1: The development framework of system.

Classification Results

Table 1: Classification results of different transformer-based models

Model	Precision	Recall	F_1-score (95% CI)
Twitter Data			
RoBERTa	0.84	**0.95**	0.89 (0.87-0.91)
SciBERT	0.87	0.89	0.88 (0.85-0.90)
BioBERT	**0.88**	0.89	0.88 (0.86-0.91)
BioClinicalBERT	0.85	0.91	0.88 (0.86-0.91)
BERTweet	**0.88**	0.91	**0.90** (0.87-0.92)
Clinical_KB_BERT	0.86	0.91	0.88 (0.85-0.90)
External: Reddit data			
RoBERTa	0.91	**0.95**	**0.93** (0.91-0.95)
BERTweet	0.89	0.90	0.90 (0.87-0.93)

- **Error analysis** show lack of context, ambiguous reference to word "migraine" as primary false positives
- Hard to spot such errors, **even for human annotator**
- Manual **Bias analysis** on 5% of all tweets in test set
- Changes in gender words didn't alter classification results

Medication Sentiments

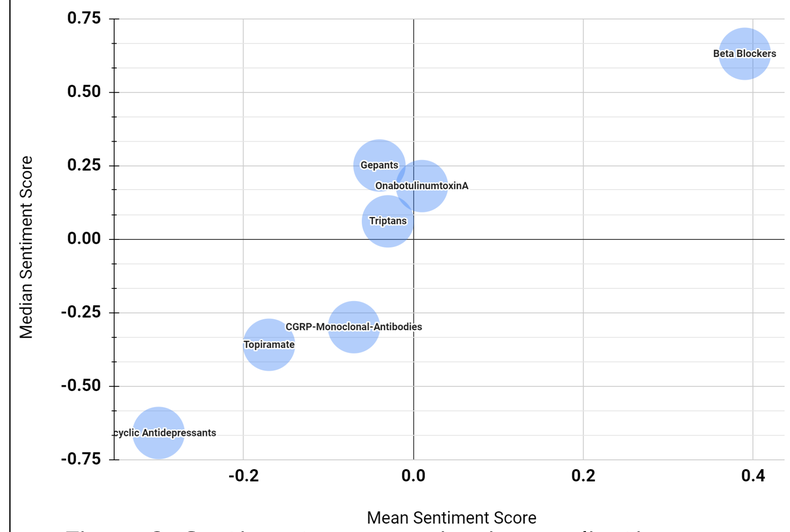


Figure 2: Sentiments across migraine medication groups

Conclusion

Social media can enhance EHRs by providing ongoing data on migraine management.

Developed NLP framework effectively analyzes social media for migraine insights.

References

1. Nittas V, et al. Electronic Patient-Generated Health Data to Facilitate Disease Prevention and Health Promotion: Scoping Review. J Med Internet Res. 2019;doi:10.2196/13320.
2. Al-Garadi MA, et al. Text Classification Models for the Automatic Detection of Nonmedical Prescription Medication Use From Social Media. BMC Medical Informatics and Decision Making. 2021, doi:10.1186/s12911-021-01394-0.
3. Ghosh S, et al. Utilizing Social Media for Identifying Drug Addiction and Recovery Intervention. IEEE International Conference on Big Data (Big Data) 20203422. doi:10.1109/BigData50022.2020.9378092.
4. Deng H, et al. Sentiment analysis of real-world migraine tweets for population research. Cephalalgia Reports. 2020;3:2515816319898867.

SCAN ME