Classification of clinical syndromes from patient-reported symptoms on social media

Stanford CS224N {Custom} Project

Evan Maestri
Department of Immunology
Stanford University
maestri@stanford.edu

Abstract

Clinical syndromes and autoimmune diseases often face diagnostic delays due to heterogeneity in patient symptoms. The goal of the project is to develop natural language processing (NLP) models that can accurately classify understudied clinical syndromes/autoimmune diseases from patient-reported symptoms on social media. The overall purpose is to mitigate diagnostic delays and reduce the dismissals of patient symptoms. By fine-tuning language models pre-trained on clinical text, the study aims to improve our understanding of chronically ill patients. Here, I examine how base model choice impacts classification performance to determine whether fine-tuning domain-specific models (Bio+ClinicalBERT) for classification performs better or similar to a domain-agnostic pre-trained BERT (DistilBERT). The main finding is that the text from patients surrounding clinical syndromes can be readily identified via fine-tuning language models, even after masking key disease terms. Both Bio+ClinicalBERT and DistilBERT perform similarly on predicting the clinical syndrome disease label. Overall, automating diagnostics in clinical syndromes may improve our ability to help these patients.

1 Key Information to include

- Mentor: Archit Sharma
- External Collaborators, sharing project: No

2 Introduction

SARS-CoV-2 has infected more than 775 million people worldwide (WHO, 2024). At least 70 million individuals around the world have long COVID (based on conservative estimated incidence of 10% of infected people) (Davis et al., 2023). With these estimates, long COVID is on track to be the third leading neurological condition within the United States. Neurological symptoms including autonomic nervous system dysfunction (dysautonomia) characterize some of the most debilitating features of post-viral syndromes (Raj et al., 2021). There is a median diagnostic delay of 2 years for dysautonomia, with only 25% of patients diagnosed within the first year of symptoms (Dysautonomia International). Additionally, some speciality clinics have >1 year wait list to be seen, with only 56 board-certified autonomic disorder specialists trying to serve a large patient population. Thus, I am interested in approaches which can automate and speed up tasks in clinical practice. In this work, I explore ways language models, especially trained in the clinical domain, could be utilized to cut down on the lag time in diagnostics for rarer and understudied conditions.

The goal of my project is to classify clinical syndromes/autoimmune diseases from patient-reported symptoms on social media. Currently there is a sparsity of NLP analysis investigating symptoms for clinical syndromes due to data availability, rarity of conditions, and much effort focused on improving acute illness outcomes. This is an important question because despite increases in prevalence of

emerging infectious diseases, the pathogenesis, diagnosis, and treatment of post-viral syndromes remain poorly understood. Often, patients with complex clinical syndromes and autoimmune diseases get wrongly dismissed that their symptoms, such as post-exertional malaise or chronic fatigue, are psychosomatic or anxiety.

I will be comparing how well multiple models are able to distinguish between: Postural Orthostatic Tachycardia Syndrome (POTS), Systemic Lupus Erythematosus (SLE), Inflammatory Bowel Disease (IBD), Ehlers-Danlos Syndromes (EDS), long COVID, and polycystic ovary syndrome (PCOS). EDS is a heritable and heterogeneous condition which impacts the connective tissues supporting joints, blood vessels, and skin. POTS is an autonomic nervous system disorder and disabling condition with patients developing extremely high heart rates (tachycardia) upon standing with symptoms including: syncope (fainting), fatigue, headaches, lightheadedness, heart palpitations, exercise intolerance, nausea, and blood pooling in extremities (Vernino et al., 2021). IBD and lupus represent two conditions which also show remarkable patient heterogeneity. PCOS is estimated to effect 1 in 10 reproductive-aged women, but up to 70% of affected women remain undiagnosed worldwide (Deswal et al., 2020).

The ability to identify complex health conditions directly from patient-reported text data using language models represents a useful investigation to improve our understanding of conditions which severely impact quality-of-life and disability. In this proof-of-concept study, I demonstrate that despite patient heterogeneity, self-reported symptoms from individuals with clinical syndromes can be utilized for disease classification.

3 Related Work

The pre-training and fine-tuning paradigm of language models has demonstrated remarkable success for many natural language processing (NLP) tasks. Releases of BERT (Devlin et al., 2019) and DistilBERT (Sanh et al., 2020) were examples of highly effective models trained largely on general domain text, such as from Wikipedia. BioBERT was also explored in the biomedical realm which was pre-trained on extensive biomedical domain corpora, including PubMed abstracts and PMC full-text articles (Lee et al., 2020). However, clinical narratives including physician notes typically have linguistic characteristics unique from general and biomedical text. Prior to Bio+ClinicalBERT there had been limited exploration of publicly available pre-trained BERT models specifically tailored to the clinical domain's specialized corpora. Bio+ClinicalBERT pre-training on clinical text data showed improvement over BioBERT or general BERT on several clinical NLP tasks by building better domain-specific language representations (Alsentzer et al., 2019).

The pre-training/fine-tuning approach also showed success on social media data as a similar application was demonstrated with COVIDTwitterBERT, which was trained on 160M tweets containing COVID-related content collected between January and April 2020 (Müller et al., 2020). Using domain-specific knowledge in the pre-training improved vaccine hesistancy predictions. Currently, many large-scale models including GPT-4 exhibit human-level performance on various tasks and academic benchmarks (OpenAI et al., 2024). This has opened the door to new possibilities of chatbot use by medical professionals and patients with great potential to improve healthcare (Lee et al., 2023).

To my knowledge, this is the first analysis of this scale where rarer/lesser studied conditions, such as POTS or EDS, have been attempted to be classified directly from text data with self-reported symptoms from social media. However, previous studies to characterize and endotype syndromes such as long COVID from patient-reported symptoms (via office visits and remote surveys) has shown success (Thaweethai et al., 2023). Inspiration for this study was drawn from using reddit/twitter posts to understand common co-occuring symptoms and outcomes in patients with long COVID (Dolatabadi et al., 2023). Analyzing the posts of self-reported symptoms better characterized our understanding of the systemic condition which affects multiple organ systems.

4 Approach

The main approach is the NLP task of multi-class text classification by assigning labels to social media posts from six diseases (Figure 1). Given the syndromes I chose (e.g. POTS, EDS) are understudied, the overall task is original and represents a useful investigation into chronic illness/disability. This represents a slightly different but related task to most electronic health record (EHR) NLP studies

attempting to identify ICD-10 codes from clinical notes (Ji et al., 2023). First, as a default baseline I demonstrate how well a prompted GPT-4 could do with this medical classification task without any fine-tuning. Next, I fine-tuned both DistilBERT (Sanh et al., 2020) and Bio+ClinicalBERT (Alsentzer et al., 2019) for the classification of the six diseases. The loss function was cross-entropy which is commonly used for multi-class classification problems. The overall goal is to have the model focus more on learning the underlying representations within the symptoms described. Thus, I wrote a custom masking function to use a masked token in place of subreddit-specific disease vocabulary. For example, during the masking process I selected terms in the top 25 most frequently occurring per subreddit which would obviously flag the model of the condition (Figure 2, Figure A.1). The terms which were masked included: ibd, uc, cd, pots, lupus, eds, heds, covid, long, pcos. The DistilBERT and Bio+ClinicalBERT fine-tuned baseline represents the upper bound on performance, and the MASK experiment will represent how well the model will do without relying on key disease-terms.

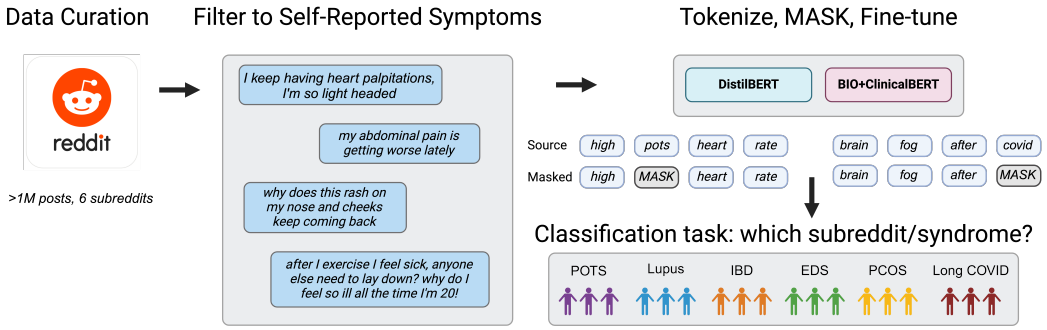

Figure 1: Custom Project Pipeline Overview

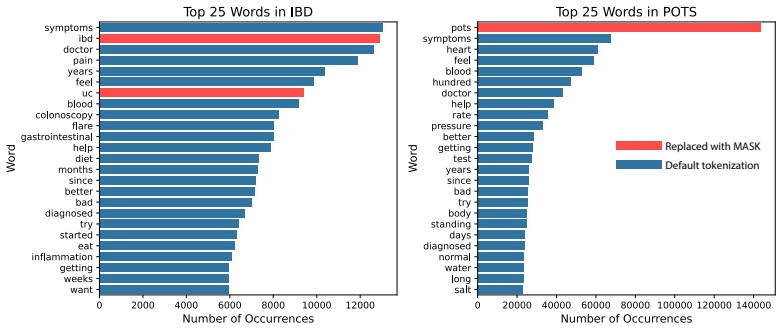

Figure 2: Top 25 words per subreddit. Red color indicates a word which was replaced with a MASK token during the masking experiment. Blue indicates words left unchanged with default tokenization.

5 Experiments

5.1 Data

Reddit data was downloaded from June 2005 to December 2023, made available largely due to PushShift (Baumgartner et al., 2020) and hosted on academictorrents.com (Watchful1 and RaiderB-Dev, 2023). Each subreddit was initially stored as zstandard compressed ndjson with the main post (submission) and comments in separate files which were later concatenated. The following subreddits were selected: POTS, IBD, ehlersdanlos, lupus, PCOS, and covidlonghaulers (Table 1). The task is to identify the source of a post from one of the six diseases.

5.2 Preprocessing

The primary interest is to curate posts which contain self-reported symptoms of the medical condition. Thus, I performed a filtration step on the raw data with regular expressions (RegEx) to capture

pronouns (e.g. I/me). For example, this includes phrases such as *in my experience, I've felt tired lately*. Next, data was cleaned using the NLTK (Bird et al., 2009) and tweet-preprocessor packages to remove hyperlinks, special characters, and punctuation. The posts were also converted to lower case and contractions were expanded. A custom mapping was defined so common abbreviations used by patients communities were expanded: ttt became tilt table test, mcas became mast cell activation syndrome. I used a cutoff minimum so the posts would have at least 50 tokens, ensuring small comments which may not include much symptom content are excluded. Next, the classes were randomly sampled to 100,000 posts per subreddit, with the exception of the smaller thread IBD (41,390 posts) which were all kept. The split of the data was 80% training, 10% validation, 10% test using stratified sampling to ensure equitable distributions of the subreddits in each split.

Subreddit	Posts, body + comments (n)	Total self-reported (n)	>50 tokens
r/POTS	507,532	357,425	176,457
r/IBD	118,652	80,254	41,390
r/ehlersdanlos	725,891	542,799	280,477
r/lupus	310,267	226,965	119,168
r/PCOS	606,227	435,283	250,308
r/covidlonghaulers	572,042	352,922	164,154

Table 1: Counts of posts per subreddit pre- and post-filtration.

Tokenization. Each post was labeled with the corresponding subreddit. The data was converted from a pandas dataframe to a huggingface dataset. Both DistilBERT (Sanh et al., 2020) and Bio+ClinicalBERT (Alsentzer et al., 2019) were utilized. A sequence maximum of 128 tokens was utilized for both models.

5.3 Evaluation method

Accuracy and macro F1 scores were used for evaluation metrics on the unseen test data as implemented in sklearn.metrics. During training the performance of individual subreddits/classes were monitored with the evaluation_report in sklearn (using F1, precision, recall). In addition to the primary test dataset ($\approx$ 10,000 posts per subreddit), a second representative test dataset (10 posts per subreddit) was evaluated for interpretability analysis using accuracy and F1.

5.4 Experimental details

For the GPT-4 baseline experiments the test dataset was fed in with the following prompt: *Please classify the text into one of the following six classes: POTS, IBD, covidlonghaulers, PCOS, ehlersdanlos, lupus. Return only one label in your output.*

To easily compare the performance of multiple models they were built using AutoModelForSequence-Classification. For both DistilBERT and Bio+ClinicalBERT finetune-baseline and finetune-MASK experiments the following parameters were kept the same to fine-tune the models and make direct comparisons: learning_rate=5e-5, num_train_epochs=4, optimizer=AdamW, batch_size=20, max_sequence_length=128, lr_scheduler_type=linear, weight_decay= 0.001, warmup_steps=100. The training time was $\approx$ two hours per model. I used early stopping if no improvements occurred in validation loss over 2 epochs to prevent over-fitting. Often, the best model was after two epochs. Using RayTune (Liaw et al., 2018) for a hyperparameter search revealed the following best parameters: num_training_epochs: 2, learning_rate: 2.75e-05, weight_decay: 0.26, lr_scheduler_type: linear, warmup_steps: 115.

5.5 Results

The performance of GPT-4 as a baseline (no fine-tuning) on the classification of the disease subreddits is indicated in Table 2. In comparison, fine-tuning DistilBERT and Bio+ClinicalBERT specifically for the classification task improved accuracy and F1 scores. For both DistilBERT and Bio+ClinicalBERT models the fine-tuned baseline experiment had accuracy and F1 around 0.83. In the MASK experiment

(where key disease terms were masked) the accuracy and F1 dropped to around 0.80. Plots of the training/validation loss and accuracy for the baseline and MASK experiments are in Figure A.2. Lastly, metrics broken down by individual subreddit (F1, precision, recall) are in Figure A.3.

Model	Accuracy	F1
GPT-4-baseline	0.638	0.666
DistilBERT-finetune-baseline	0.831	0.833
Bio+Clinical BERT-finetune-baseline	0.832	0.835
DistilBERT-finetune-MASK	0.807	0.808
Bio+Clinical BERT-finetune-MASK	0.805	0.806

Table 2: Comparison of model performance results on unseen test data.

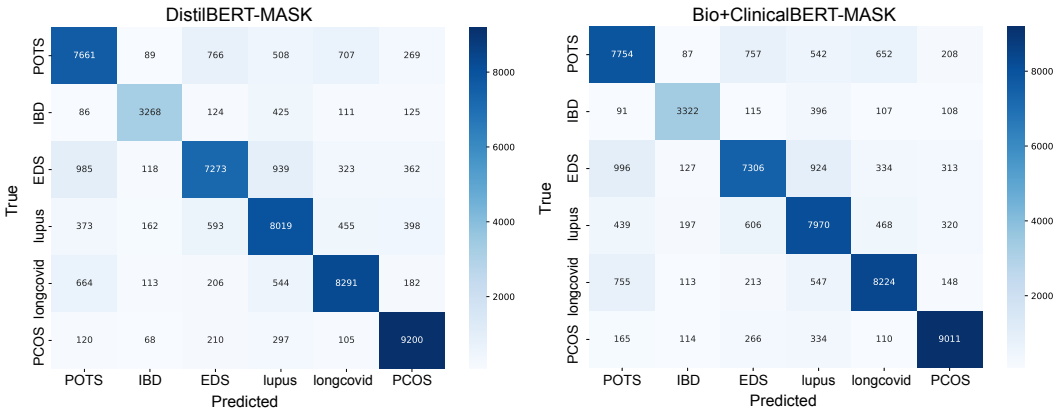

Figure 3: The confusion matrix from the test set for both models from the MASK experiment.

Given I did not use any metrics to access social media post quality (aside from a threshold of >50 tokens) and many posts in the test set could still contain noise (e.g. little to no symptoms described) these results performed better than I anticipated. Overall, the result trends were expected though because I first show how GPT-4 as a general purpose language model performs on the task. It would understand most terminology, but perhaps not all precise medical disease-specific terminology which would be helpful for the classification. Then, when I explicitly fine-tune on the subreddit-specific symptomatology, which may include precise terms and linguistic characteristics utilized by patient communities, this boosts the performance. Lastly, by masking out top key words, it makes sense that the performance would drop once the model no longer relies on them and tries to focuses on relationships in the symptoms more. Initially, I hypothesized Bio+ClinicalBERT would outperform the domain-agnostic BERT. However, both models could be performing similarly for this task because the social media posts (despite having clinical information) may not resemble the semantic structure of clinical notes enough to mount high benefit from domain-specific pre-training here.

6 Analysis

I investigated how many posts would be needed to achieve adequate classification performance given high quantities of data do not exist for many rarer conditions. For the dataset size analysis the same parameters were selected as above using DistilBERT-finetune-MASK, however, the number of posts per subreddit in the training were gradually increased from 50 to 100k posts. The test set size was always 5,000 posts. An F1 score of 0.75 was achieved at around 25k posts, which may help inform future iterations in data collection.

The overarching goal of the project was to build models which learn from the symptoms. In order to test that the model actually relies on symptoms and examine when it succeeds and when it falls

short, I curated a subset of 10 example posts per subreddit. For GPT-4, the accuracy was 0.666 and F1 0.638. For DistilBERT-finetune-MASK the accuracy on the intepretability subset was 0.9 and F1 0.89. GPT-4 had a harder time with examples for POTS, covidlonghaulers, and lupus. Many of the POTS curated example posts were misclassified by GPT-4 as ehlersdanlos, but were properly classfied by the fine-tuned DistilBERT.

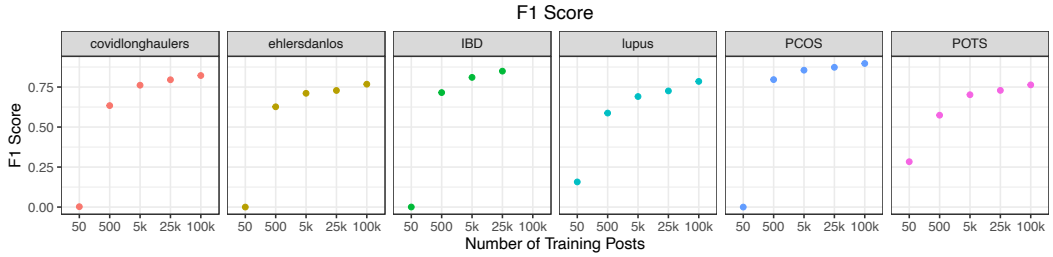

Figure 4: Performance with varying dataset training sizes.

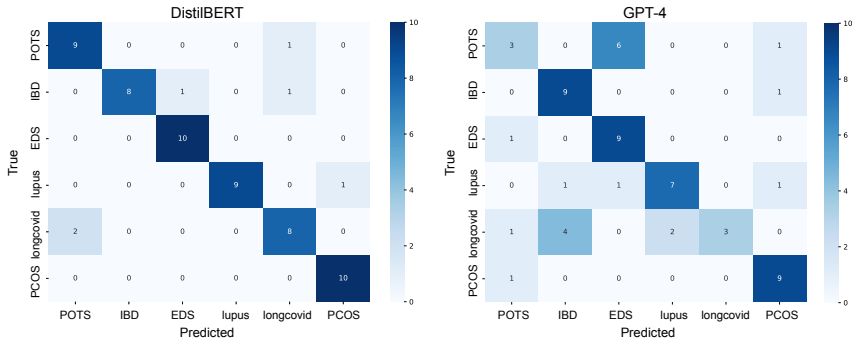

Figure 5: Confusion matrix on the subset of posts for interpretability analysis.

The following section will now be based on the DistilBERT interpretability results. Many well-characterized disease symptoms were accurately learned by the model, such as the following:

POTS

- *I experience dizziness and fainting frequently.*
- *I have high heart rates and blood pooling in my extremities.*

covidlonghaulers

- *I am near six months past initial infection with lingering cough chest pain and fatigue.*
- *I cannot even remember how my brain worked before this. I felt a sudden change my thoughts became slow, my memory fails me now.*

lupus

- *I have a butterfly rash and see a rheumatologist.*
- *I think this photosensitivity is causing my skin to flare.*

ehlersdanlos

- *A trainee asked if I was flexible today and I bent my thumb back to my forearm and said yup.*
- *Someone please explain why I bruise so easily and I have such stretchy skin.*

All examples which were used for the interpretability analysis can be found in Table A.1.

Using example inputs which just list medications commonly used by that community get classified correctly; for example *propranolol/corlanor* for POTS or *mycophenolate mofetil/plaquenil* for lupus. Similarly, the model classifies accurately based on phrases for common tests for diseases such as *dsDNA/ANA blood tests* which are common screens for suspected lupus or the *Beighton score* (a test for joint hypermobility) which is used for EDS.

I also gave a few ambiguous phrases which could fit in more than one category *I have chronic fatigue, headaches, and feel like I can't workout* was predicted as POTS, despite my initial covidlonghaulers label. Similarly, one I intended for POTS was predicted as covidlonghaulers *after trying to do any exercise I usually have problems with thinking memory and concentration*. The phrase *I suffer from shortness of breath* was an intended covidlonghaulers label, but was predicted as POTS. Lastly, the phrase *I have persistent gastrointestinal symptoms and my digestion is off after I eat.* which was intended for IBD got classified as covidlonghaulers. This could be due to the phrase "viral peristence" being common in the covidlonghaulers subreddit, due to the high prevalence of digestion issues in long COVID, or because of the class imbalance (less IBD training posts than covidlonghaulers).

A limitation of the current model is that for PCOS due to the high subreddit occurrences of the word *weight*, (where weight gain is a common symptom) using that word alone appears to get classified into the PCOS category. For example, the phrase *I'm not happy with my weight* is more general (not PCOS-specific) but gets classified as PCOS. In future iterations, I may consider masking the word weight, since PCOS also had the highest F1 of all six subreddits.

7 Conclusion

In this work, I have shown that by fine-tuning language models using reddit data of patient-reported symptoms, I can accurately classify clinical syndromes from text. Consistently some of the most common words used across the clinical syndrome subreddits include: *help* and *pain*. Patients in these communities express strong sentiments demonstrating they are tired of being neglected with a clear plea for better healthcare, which language models may be able to solve. The primary finding is that both DistilBERT and Bio+ClinicalBERT achieve an accuracy of $\approx 80\%$ for a six-disease classification task even after masking key disease-specific terms. In interpretability analysis I determined that well-characterized disease symptoms, medications, and clinical tests for certain syndromes are understood by the NLP models.

I think that the current approach (one post corresponding to one condition) may struggle with the fact that clinical syndromes can be nebulous; one individual may experience multiple constellations of symptoms at once which can fit in more than one clinical syndrome category. For example, EDS and POTS are often comorbidities. Many individuals with long COVID also likely have POTS or some form of autonomic nervous system dysfunction. The underlying assumption of the present study is that a user posting in one subreddit (e.g. IBD) is likely to be discussing their primary symptoms related to the subreddit. Future directions include modifying the architecture and training process to allow one post to have multiple labels. The multi-label approach would enable a better understanding of the overlapping nature of symptoms and comorbidities in clinical syndromes.

Given my interest in diagnosing chronic diseases and clinical syndromes, a model like Bio+ClinicalBERT which instead was trained on clinical notes from non-ICU visits may be useful. This was a proof-of-concept study to demonstrate signal in symptoms from the text data of patients with clinical syndromes. Future versions will instead build NLP models using EHR clinical notes, which will ensure the training data has physician-confirmed labels. NLP models which try to diagnose patients from text in the EHR realm may face challenges in curating enough medical records for rare syndromes, to ensure the significant differences in care practices across institutions are accounted for.

8 Ethics Statement

A method which could flag text from patients as potential for POTS, PCOS, EDS, etc. could face ethical concerns if an improper classification or predicted diagnosis occurred. The use of this model should act as an initial triage step to guide patients towards the most appropriate physician who may cut down on the lag time in the actual diagnosis. It should not be positioned as a replacement for clinical diagnosis but rather as a tool to assist users in seeking appropriate medical care. One limitation is that individuals posting in these group social media channels may be seeking help for a

potential diagnosis since clinical syndromes can take a long time to receive adequate care. The labels are not clinician diagnosis confirmed, which would require additional testing including a tilt table test for POTS or positive blood work showing specific biomarkers for lupus. Due to the complexity of clinical syndromes, use of this type of model should not be replaced by clinical diagnosis from a healthcare professional, and individuals should consult with physicians for proper evaluation of any suspected condition.

As a mitigation strategy, educational resources could be provided alongside the predictions to help users understand model limitations and determine which healthcare specialty (rheumatology, gastroenterology, cardiology, neurology, etc.) may be appropriate to consult. Another concern would be for companies or other entities to download a users social media data and get access to what health conditions the individual might have leading to discriminatory practices against those with chronic diseases and disabilities (e.g. in hiring or insurance claims). Limiting the model scope and use (e.g. the number of classifications per time period) could prevent mass download and potential misuse.

References

Emily Alsentzer, John R. Murphy, Willie Boag, Wei-Hung Weng, Di Jin, Tristan Naumann, and Matthew B. A. McDermott. 2019. Publicly Available Clinical BERT Embeddings. In *arXiv preprint arXiv:1904.03323.*

Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. The Pushshift Reddit Dataset. In *arXiv preprint arXiv:2001.08435.*

Steven Bird, Ewan Klein, and Edward Loper. 2009. Natural Language Processing With Python: Analyzing Text With The Natural Language Toolkit. O'Reilly Media, Inc.

Hannah E Davis, Lisa McCorkell, Julia Moore Vogel, and Eric J Topol. 2023. Long COVID: Major findings, mechanisms and recommendations. *Nat. Rev. Microbiol.*, 21(3):133–146.

Ritu Deswal, Vinay Narwal, Amita Dang, and Chandra S Pundir. 2020. The prevalence of polycystic ovary syndrome: A brief systematic review. *J. Hum. Reprod. Sci.*, 13(4):261–271.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training Of Deep Bidirectional Transformers For Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Elham Dolatabadi, Diana Moyano, Michael Bales, and et al. 2023. Using Social Media To Help Understand Patient-Reported Health Outcomes Of Post-COVID-19 Condition: Natural language Processing Approach. *J. Med. Internet Res.*, 25:e45767.

Shaoxiong Ji, Wei Sun, Xiaobo Li, Hang Dong, Ara Taalas, Yijia Zhang, Honghan Wu, Esa Pitkänen, and Pekka Marttinen. 2023. A Unified Review of Deep Learning for Automated Medical Coding. In *arXiv preprint arXiv:2201.02797.*

Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2020. BioBERT: A Pre-Trained Biomedical Language Representation Model For Biomedical Text Mining. *Bioinformatics*, 36(4):1234–1240.

Peter Lee, Sebastien Bubeck, and Joseph Petro. 2023. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *N. Engl. J. Med.*, 388(13):1233–1239.

Richard Liaw, Eric Liang, Robert Nishihara, Philipp Moritz, Joseph E Gonzalez, and Ion Stoica. 2018. Tune: A Research Platform for Distributed Model Selection and Training. *arXiv preprint arXiv:1807.05118.*

Martin Müller, Marcel Salathé, and Per E Kummervold. 2020. COVID-Twitter-BERT: A Natural Language Processing Model To Analyse COVID-19 Content On Twitter. In *arXiv preprint arXiv:2005.07503v1.*

OpenAI, Josh Achiam, Steven Adler, and et al. 2024. GPT-4 Technical Report. In *arXiv preprint arXiv:2303.08774.*

Satish R Raj, Amy C Arnold, Alexandru Barboi, and et al. 2021. Long-COVID Postural Tachycardia Syndrome: An American Autonomic Society Statement. *Clin. Auton. Res.*, 31(3):365–368.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2020. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. In *arXiv preprint arXiv:1910.01108.*

Tanayott Thaweethai, Sarah E Jolley, and RECOVER Consortium et. al. 2023. Development Of A Definition Of Postacute Sequelae Of SARS-CoV-2 Infection. *JAMA*, 329(22):1934–1946.

Steven Vernino, Kate M Bourne, Lauren E Stiles, and et al. 2021. Postural Orthostatic Tachycardia Syndrome (POTS): State Of The Science And Clinical Care From A 2019 National Institutes of Health Expert Consensus Meeting - Part 1. *Auton. Neurosci.*, 235(102828):102828.

Watchful1 and RaiderBDev. 2023. Reddit comments/submissions 2005-06 to 2023-12. Online.

World Health Organization: WHO. 2024. WHO Coronavirus (COVID-19) Dashboard.

A Appendix

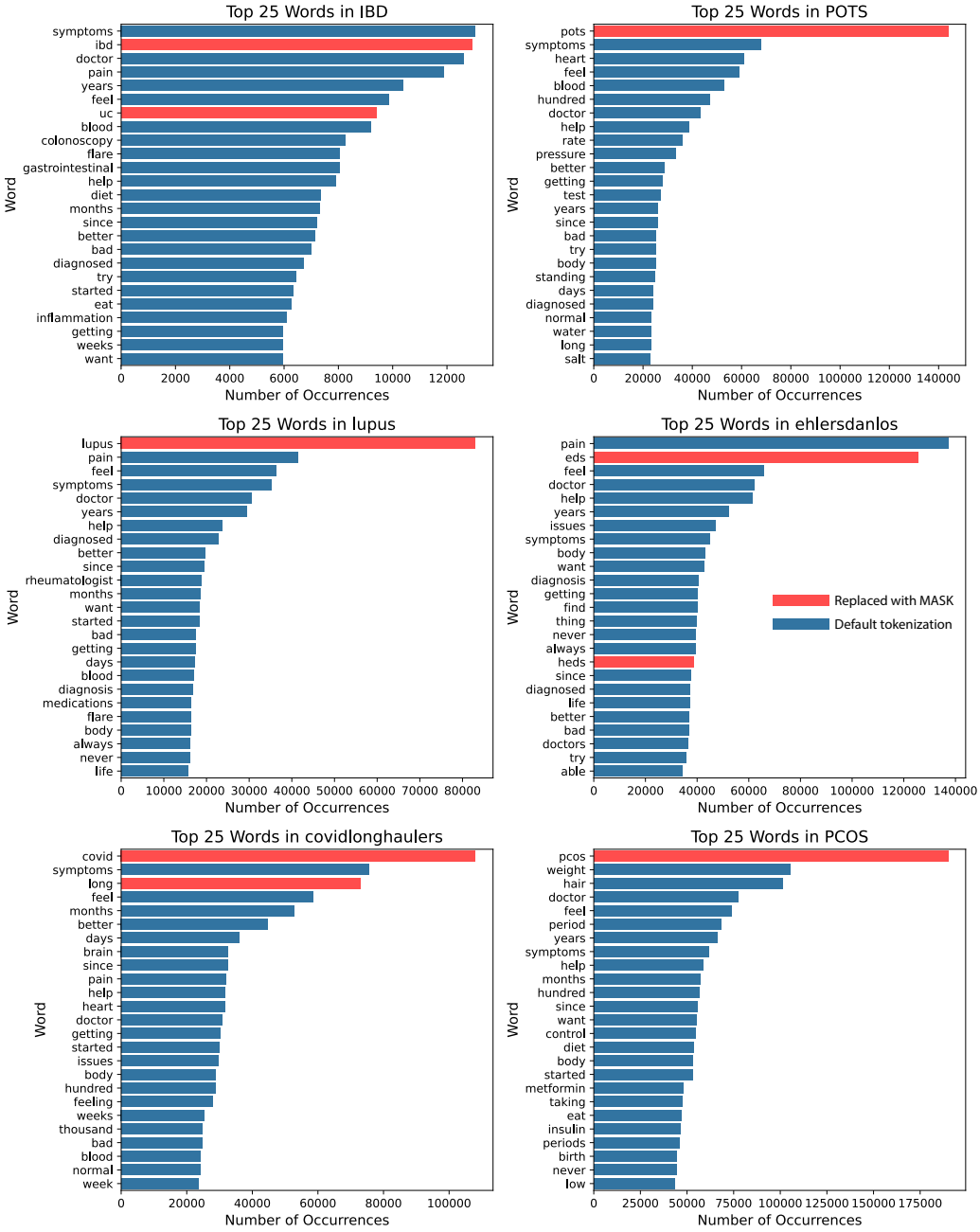

Figure A.1: Top 25 words per subreddit. Red color indicates a word which was replaced with a MASK token during the masking experiment. Blue indicates words left unchanged with default tokenization.

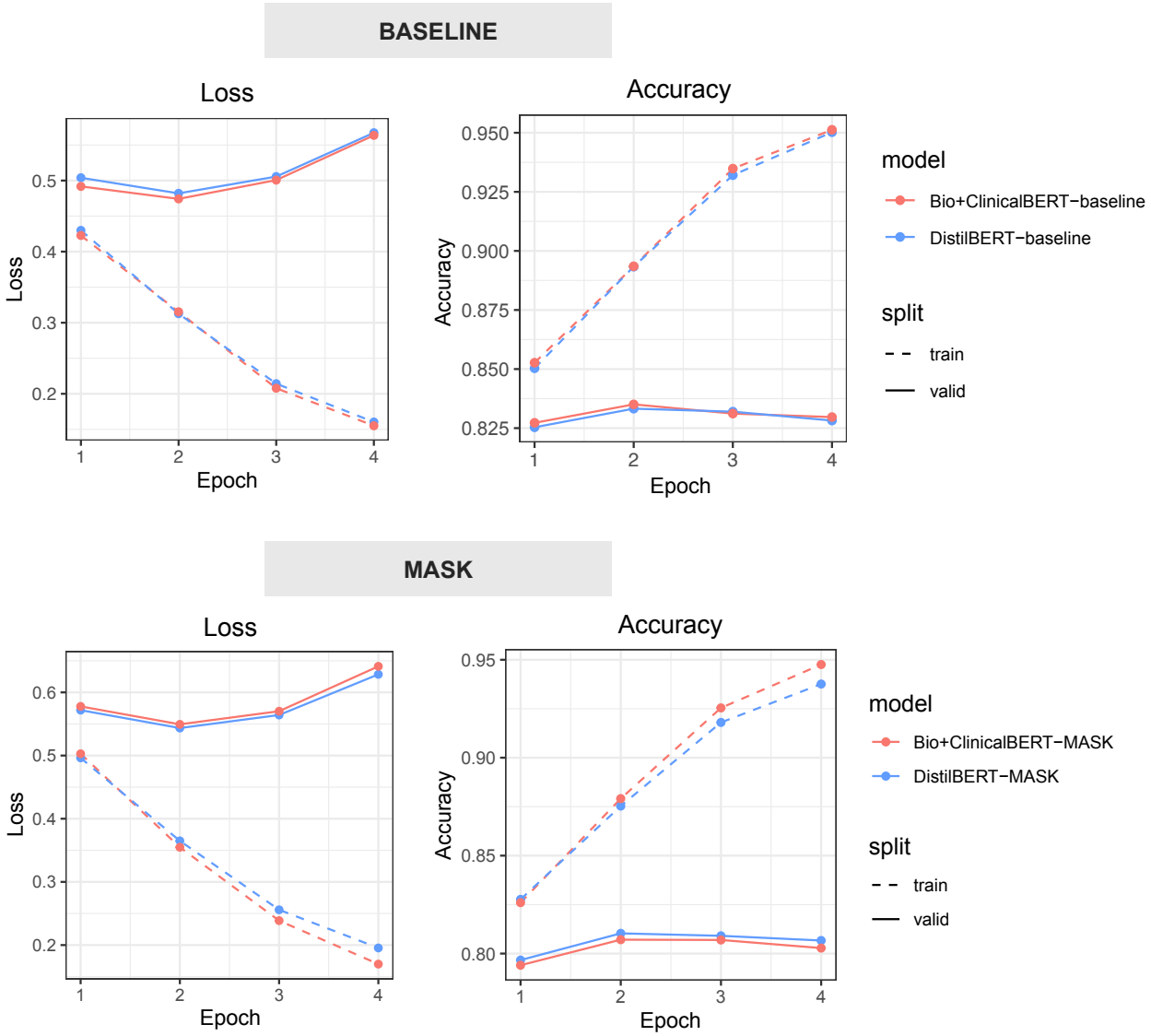

Figure A.2: Training and validation loss and accuracy for both models at baseline (top) and after masking key disease terms (bottom). The dotted line is training and solid line is validation. Plots are colored by model with DistilBERT as blue and Bio+ClinicalBERT as red.

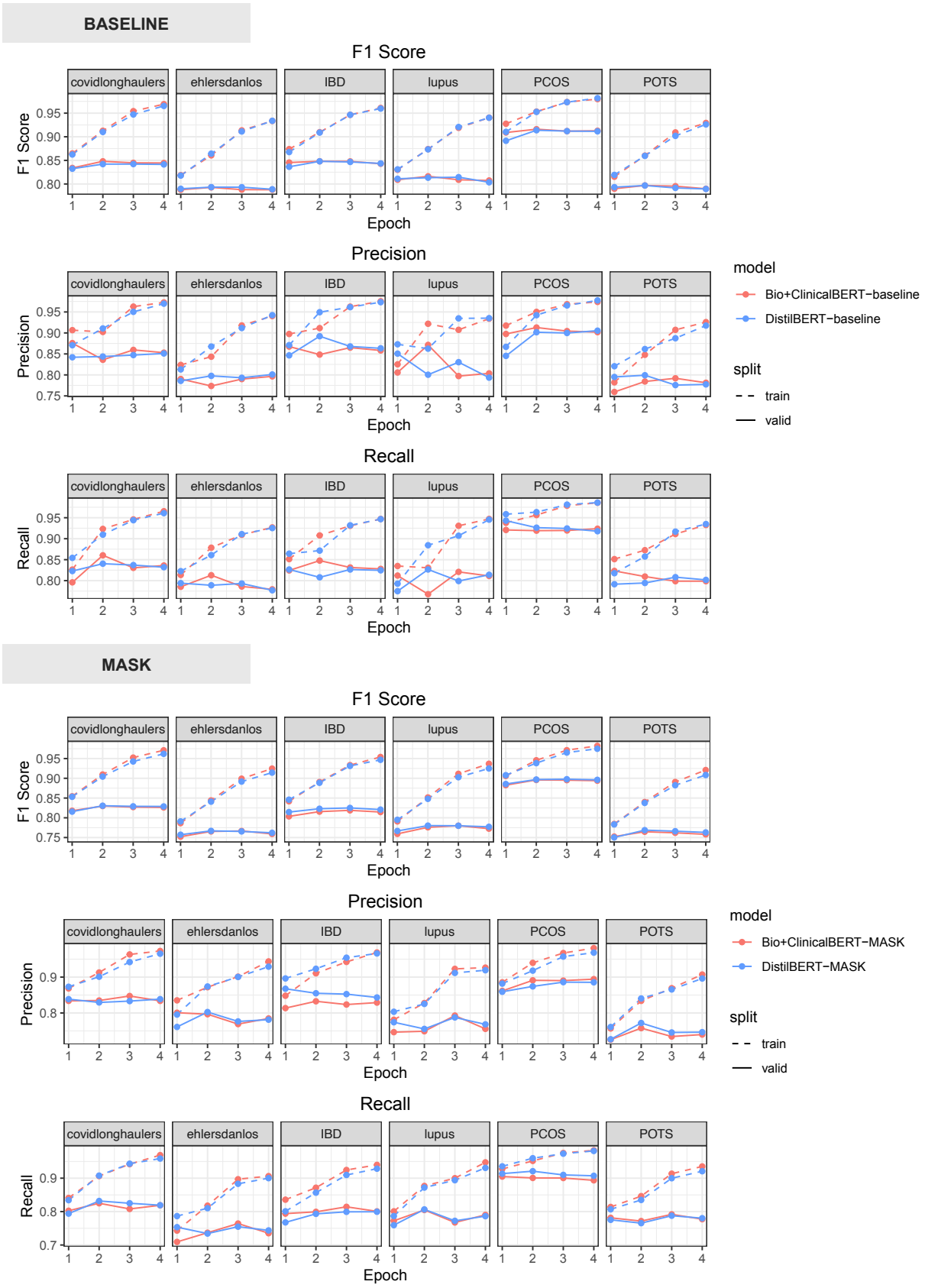

Figure A.3: Metrics tracked per subreddit during training for the baseline and the masking experiment.

Table A.1. Example posts per subreddit for interpretability analysis.

POTS

- I experience dizziness and fainting frequently.
- I have high heart rates and blood pooling in my extremities.
- I do not tolerate upright positions well even after sitting for a very long time.
- I find increasing salt and water intake, compression socks, and eating smaller meals helps my symptoms.
- Does anyone else get this purple discoloration in your hands and feet, like the veins? Should I be worried?
- After trying to do any exercise, I usually have problems with thinking, memory, and concentration.
- I take propranolol.
- I take midodrine in addition to corlanor which helps my near syncope.
- I have headaches, low blood volume, and issues with small fiber nerves in my legs.
- I feel really awful in the heat, even though I drink electrolytes. I get breathless and sweaty easily.

covidlonghaulers

- I have chronic fatigue, headaches, and feel like I can't workout.
- I suffer from shortness of breath.
- I am near six months past initial infection with lingering cough, chest pain, and fatigue.
- Mild fever ebbing and flowing, never above ninety-nine, with a sore throat and PEM.
- My main symptoms I did have were chest pain, brain fog, and digestion issues.
- Inflamed whenever I breathe in, I feel pressure, it is difficult to breathe.
- I'm having poor concentration, general weakness, and I feel tired all the time.
- Ugh, I still have occasional chest pain on exercise and I seem to need more rest than usual.
- I cannot even remember how my brain worked before this. I felt a sudden change my thoughts became slow, my memory fails me now.
- I am getting non-restorative sleep.

lupus

- I have a butterfly rash and see a rheumatologist.
- I take mycophenolate mofetil and I tried plaquenil.
- I think this photosensitivity is causing my skin to flare.
- Recent blood work showed a positive dsDNA.
- I tested positive for ANA, what does this mean?
- Hair loss has been the worst part for me. I can deal with the pain, but the embarrassment of hair loss is so depressing.
- I have also had weird skin things, scalp sores, and mouth sores going on, maybe some Raynaud or heat intolerance.
- My doctor said I might have proteinuria, so something is affecting my kidneys.
- Kept having these flares though, and my face would swell up.
- Luckily, I responded well to prednisone, hydroxychloroquine and now my health is better managed.

IBD

- I have persistent gastrointestinal symptoms and my digestion is off after I eat.
- I ate pizza after work and boy did I regret it with cramps, diarrhoea, and some constipation.
- Dairy, specifically milk, really does me in. Sugar or spicy food does not do anything to me, but caffeine makes me have horrible cramps.
- The preparations for a colonoscopy is rough.
- I would have to rush to the bathroom or take trip after trip.
- A pain in my lower stomach every time before I go.
- I get a feeling like I never feel like I have emptied out after pooping.
- I have delayed gastric emptying and my gut motility always feels off.
- This flare has included dark blood in my stool.
- I am scared to leave the house and especially eat without the safety of the bathroom near me.

ehlersdanlos

- I experience joint pain and sometimes I dislocate my shoulder.
- A trainee asked if I was flexible today and I bent my thumb back to my forearm and said, yup.
- I have cervical instability.
- Ugh, it is literally the worst. My hip popped out of place yesterday while standing completely still in my kitchen.
- I keep having subluxations where I go in and out of phases of chronic joint pain for varying reasons.
- I went to a PT for knee pain and they made sure I did not overstretch.
- I sleep with pillows under my knees to keep them from hyperextending in my sleep.
- Was screened with the Beighton score today.
- Someone please explain why I bruise so easily and I have such stretchy skin.
- Always been flexible through life. My shoulders popped out always in childhood with extreme muscle pain after physical activity.

PCOS

- I have had excessive weight gain in the last year.
- I am twenty-one and I have spent the whole year trying to lose the weight. Desperate for help I made a doctor's appointment.
- I am not happy with my weight.
- I have high cortisol stress hormones.
- Now it is pretty bad my periods are irregular and I have not had one since late December.
- I am still having issues with acne and hair growth.
- Increased hair growth on my face and body and balding at the top of my head.
- My belly always looks pregnant and no matter what I do, it does not go down.
- My bloodwork came back normal, but I keep having irregular periods and infertility issues.
- Weight.