Covid-19 Vaccine Tweet Annotation
Using Bidirectional Encoder Representations from Transformers

Derek Carey
Derek.carey@outlook.com

Abstract

Vaccine hesitancy represents a significant risk in global and regional efforts to control and eventually end the ongoing worldwide COVID-19 pandemic. With the high volume of social media activity discussing vaccination efforts, modern natural language processing techniques can allow real-time monitoring of public sentiment regarding COVID-19 vaccinations. To measure sentiment, Tweets from April 15th, 2021 to May 21st, 2021 were collected and manually annotated for assessment of author vaccination status, side-effect reference, vaccine sentiment, and negative sentiment reason. Using this labeled dataset of 2000 Tweets, several Bidirectional Encoder Representations from Transformers (BERT) models were fine-tuned to predict labels for each annotation task. Measuring model accuracy, a fine-tuned version of the Electra_base BERT model implemented with uncased text normalization produced the best classification results. Electra_base achieved 95% accuracy in determining author vaccination status and 89.8% accuracy in side-effect reference classification measured on test dataset Tweets. Testing also demonstrated 79.8% accuracy in sentiment analysis and 85.8% accuracy in determining a reason for a Tweet's negative sentiment. Results showcase the ease of developing accurate annotation models using transfer learning. With additional refinement, classification models are ready for use in real-time monitoring of public vaccination sentiment on Twitter or other social media platforms.

Introduction

The rapid spread of COVID-19 in early 2020 caused an immediate economic and cultural response. Aided by significant investment from world governments and pharmaceutical companies, medical researchers conducted testing and clinical trials for over 80 vaccines. By

February 27th, the United States F.D.A had issued emergency use authorizations for three vaccines developed and manufactured by Johnson & Johnson, Pfizer-BioNTech, and Moderna. (Zimmer et al., 2020)

Public trust in vaccine research and development processes has a significant impact on intention to receive COVID-19 vaccinations. From a March 2021 survey of Americans, Boston Consulting Group estimated about 80 million adults would be hesitant to take any vaccine if it were available to them at no cost. Social media platforms such as Facebook and Twitter are a source of news and information for millions of people and serve as public discussion forums. Data from these services can provide significant insight into public trust in vaccination efforts. However, the unstructured nature and volume of this data make meaningful interpretation a challenge. At-scale manual annotation is impractical beyond several thousand tweets. Application of machine learning, deep learning, and transfer learning techniques assists with the task of automating annotation of COVID-19 vaccine-related Tweets for multiple sentiment analysis and classification tasks. Accurate annotation allows rapid assessment of prevalence vaccine hesitancy in specific geographic regions, and this data can assist in initiatives to improve social media discourse about vaccination.

Literature Review

The development and administration of vaccinations have been components of campaigns to combat infections since the 1800s. The use of vaccinations as instruments of government expanded in the mid-20th century with the foundation of the World Health Organization (WHO) in 1948. In 1966, the WHO announced a vaccination program with the goal of globally eradicating smallpox. (Blume et al., 2017, 6) This was achieved in 1980 when the WHO certified the world was free of naturally occurring smallpox. Successfully eliminating smallpox required considerable leverage of political will with extensive efforts expended to administer vaccines in high-outbreak regions in India and Southeast Asia.

The relationship between politics and vaccination has historically led to conflict between governments attempting to implement vaccination policies and the public. In 1998, British surgeon Andrew Wakefield published a paper to the journal *The Lancet* suggesting a link between the triple vaccine for measles, mumps, and rubella (MMR) and the development of autism. (Blume et al., 2017, chap. 9) In 2010, *The Lancet* retracted the publication after the British General Medical Council determined falsification of research from Wakefield. However, Wakefield's research had already been subject to significant attention in British media. Government response to the Bovine spongiform encephalopathy (BSE) epidemic in the 1980s and 1990s eroded the public trust in science and public health. Left-leaning newspaper *The Guardian* initially took a critical response to Wakefield's research in 1998. Following the birth of prime minister Tony Blair's son in 2000, *The Guardian* began publishing opinion pieces critical of Tony Blair's refusal to state whether his son received the MMR vaccine. Tabloids such as the *Daily Mail* continued to criticize the MMR vaccine even after a 2005 Cochrane review finding no significant association between MMR immunization and autism as well as other childhood disorders. Contemporary public vaccination sentiment continues to be influenced by these public debates surrounding the MMR vaccine over two decades later.

In 2019 the WHO listed vaccine hesitancy as one of ten global health threats amidst a 30% rise in global measles cases. Defining vaccine hesitancy as “the reluctance or refusal to vaccinate despite the availability of vaccines,” the WHO estimated 1.5 million deaths could be avoided annually with improvements in global vaccination coverage. Growth in social media usage over the 21st century has shifted where the public conversation about vaccinations took place. Hilary Piedrahita-Valdés et al. (2021) analyzed 1,295,823 English and 203,404 Spanish Tweets from 2011 to 2019 for vaccine relation and sentiment. While there were only 2700 vaccine-related Tweets in 2011, a total of 57,667 vaccine-related Tweets were recorded in April of 2019. Using a support vector machines (SVM) based classification model with > 85% accuracy, they determined that approximately 69.37% of Tweets were neutral, 21.78% Tweets were positive, and 8.86% of Tweets were negative in sentiment.

Following significant spread, the WHO declared COVID-19 a global pandemic in March 2020. As vaccination remains the most effective solution in reducing the spread of the virus, COVID-19 vaccination is a common topic of discussion on social media. Public opinion and even the language used to regard COVID-19 vaccination is constantly evolving, and previous research on general vaccination sentiment may be less relevant when assessing COVID-19 vaccine hesitancy. Training deep learning classification models on manually annotated COVID-19 related Tweets provides a solution for real-time monitoring of changes in online vaccination sentiment.

Data collection and annotation schema strategies for COVID-19 tweet annotation can be modeled off previous research in vaccine sentiment analysis. A 2017 study on HPV vaccine sentiment published in the *Journal of Biomedical Semantics* by Du et al. (2017) utilized a multi-step annotation structure to manually create a dataset of 6000 tweets. Annotators first determined whether a tweet was related or unrelated to HPV vaccination. Related tweets were classified by positive, negative, or neutral sentiment. Negative sentiment tweets were further annotated for a specific reason. These reasons consisted of safety, efficacy, cost, cultural or emotional, or other concerns with HPV vaccination.

Li Zhang et al. (2020) utilized the same HPV dataset to assess the efficacy of transfer learning for vaccine sentiment annotation by fine-tuning Bidirectional Encoder Representations from Transformers (BERT) and generative pre-training (GPT) models. Their fined-tuned BERT model achieved a Micro-F1 score of 0.769, representing a 6.95% improvement over the SVM-based machine-learning model used by Du et al. BERT makes use of the Transformer described in the paper "Attention Is All You Need." (Vaswani et al., 2017). Transformers include an attention mechanism that are capable of learning contextual relations between words within a text. The original Transformer proposed by Vaswani et al. consists of an encoder that reads the text input and a decoder that produces a prediction for the given task such as converting English text to French. BERT only requires the encoder mechanism for generating a language model capable of contextual knowledge. Unlike traditional left-to-right or right-to-left language models, BERT utilizes a "masked-language model" or MLM to perform bidirectional pre-training on unlabeled text. (Devlin et al., 2019) BERT is simultaneously pre-trained for a binarized next sentence prediction task, assessing whether a provided sentence is the correct proceeding sentence or of no immediate relation to a preceding sentence. BERT-Base, one of several versions of the original architecture, consists of 12 hidden layers, 768 hidden sizes, and 12 attention heads with over 110 million parameters. Pre-trained using the BooksCorpus (Zhu et al., 2015) and English Wikipedia, BERT-Base achieved a 79.6 average score against the General Language Understanding Evaluation (GLUE) benchmark. (Wang et al., 2018a) The GLUE benchmark incorporates performance tests against a range of natural language tasks, including sentence-pair completion and question answering.

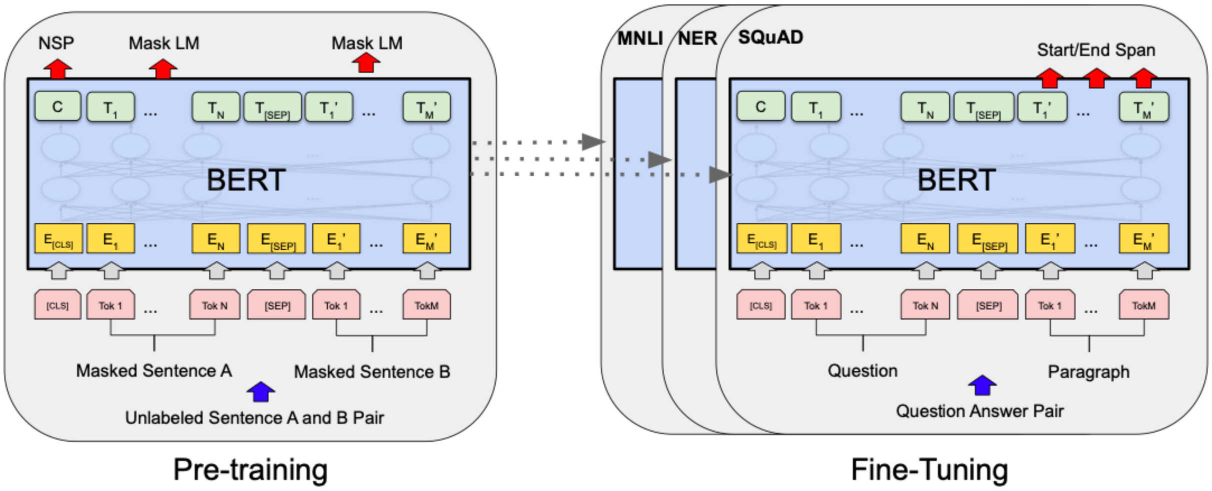

Figure 1. BERT pre-training and fine-tuning procedures as described in “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” (Devlin et al., 2019)

There are numerous notable bi-directional language representation models that emerged following the original BERT publication. "Efficiently Learning an Encoder that Classifies Token Replacements Accurately", or ELECTRA, modifies the token masking process used in BERT. (Clark et al., 2020) Instead of replacing some tokens with [MASK], ELECTRA is pre-trained by replacing some tokens with plausible alternatives sampled from a generator. Clark et al. modified BERT’s pre-training task from masked word identification to predicting whether a generator has replaced a token within a word sequence. ELECTRA-Base++, trained using comparable same parameter settings to BERT-Base on Wikipedia+BooksCorpus dataset, achieved an average GLUE score of 83.5. ELECTRA also demonstrated strong performance against GLUE benchmarks for small models. Other notable models include RoBERTa, which was developed by extensively measuring adjustments to hyperparameters to optimize BERT pre-training. (Liu et al., 2019) Bi-directional language representation models remain an active area of research and innovation.

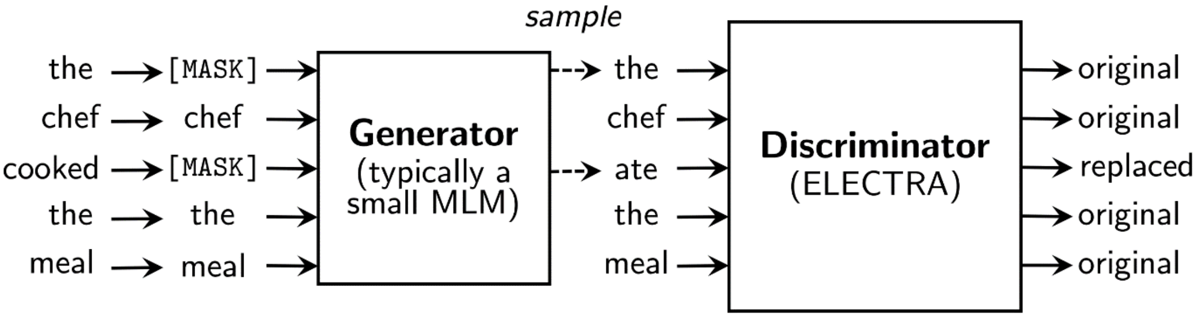

Figure 2. ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators (Clark et al., 2020)

BERT and other Transformer encoder architectures have proven extraordinarily successful in various natural language processing tasks. They are freely available to download and fine-tune using deep learning packages such as Tensorflow. Performance and ease of implementation were significant factors that led to choosing to fine-tune BERT and Electra for developing COVID-19 Tweet annotation models.

Data

Tweets related to COVID-19 vaccination were collected between April 15th and May 21st, 2021, using the Twitter standard API. API requests used applied filters for English and matched against keywords “covid” and “vaccine”. ReTweets, Tweets with mentions, and Tweets with media and URLs were excluded from data collection by modifying the Twitter API request query. Following assembly, 2000 Tweets were annotated by hand with attempts to exclude any Tweets suspected of being from a non-personal Twitter account from annotation. Tweets were evaluated to ensure content related to COVID-19 vaccinations. Research objectives were not structured around categorizing Tweets by relation or non-relation to the COVID-19 vaccine, so any non-related Tweets were excluded from the corpus.

After establishing the corpus, an assessment was made on whether each Tweet explicitly stated or implied the author received a vaccination or had a pending appointment. Tweets expressing positive sentiment for vaccination where the author's vaccination status was unknown were marked as a 'No.' Tweets were evaluated on whether the author referred to side-effects associated with COVID-19 vaccines regardless of vaccination status or sentiment. For example, a Tweet referencing arm soreness after a second vaccination and a Tweet where the author states they will not get a vaccine due to "covid vaccination deaths" both meet the criteria for side-effect reference.

An annotation was made based on whether a Tweet expressed positive, negative, or neutral sentiment towards vaccination. Vaccine sentiment is subjective, and it was difficult to establish strict definitions on what is positive, negative, or neutral. As a guideline, Tweets expressing explicit enthusiasm towards vaccination or criticism of anti-vaccination were annotated as positive in vaccine sentiment. Comments disparaging anti-vaccination were annotated as positive in sentiment even if the language used was inflammatory. Tweets lacking in any explicit emotion towards vaccination were neutral regardless of whether the author had received a vaccination. For example, "I got my covid vaccination" would be considered a neutral sentiment Tweet.

Tweets classified as negative in sentiment either discouraged vaccination efforts or expressed frustration in vaccination efforts and administration. Negative sentiment Tweets were subcategorized based on common reasons, which included concerns about side effects, efficacy, availability, or 'other.' Tweets labeled 'other' might express negative cultural, political, or emotional negativity towards vaccination. Tweets with a positive or neutral sentiment received a reason of 'Not Negative.' Only the dominant sentiment reason received the label even in cases where Tweets expressed multiple concerns.

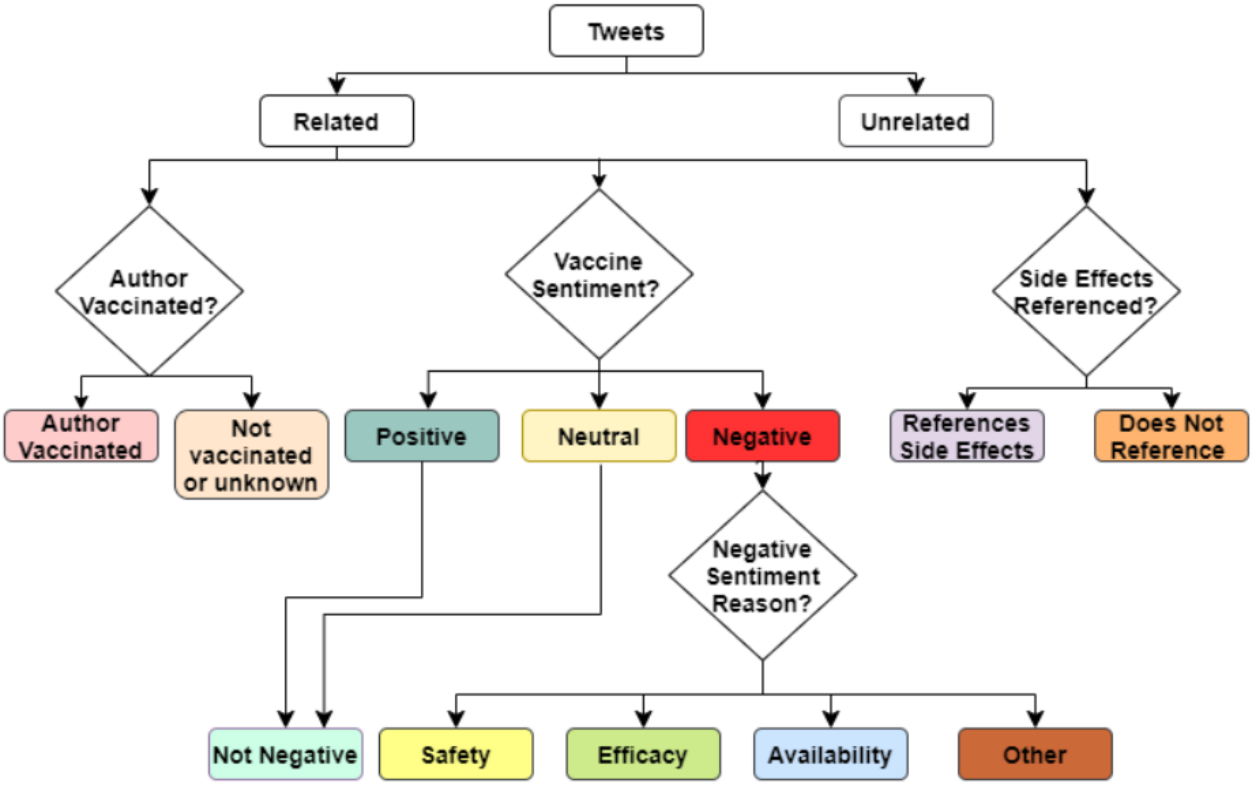

Figure 3. Sentiment classification scheme for COVID-19 vaccine related Tweets. The categories in colored boxes are all possible labels that can be assigned to a Tweet.

Corpus documents underwent several data preparation steps before model training. Emojis are helpful when manually assessing the sentiment of a Tweet but remain a challenge for natural language processing. A function converted Unicode emoji characters present in corpus text to CLDR short names. Specialized models downloaded from TensorFlowHub performed any additional text preprocessing. These preprocessing models perform text normalization of corpus data suitable for each pre-trained model, such as stripping accent markers or adjusting text casing before tokenization into word pieces. Shuffling the corpus using a random seed allowed splitting it into the separate train, test, and validation datasets. The training dataset consists of 70% of total records, while the test and validation datasets make up 20% and 10% of corpus records. Once shuffled, train, validation, and test datasets were kept consistent for all model training and evaluation tasks to allow for a fair comparison. Each dataset was inspected to ensure the proportions of each Tweet label were comparable to proportions in the complete corpus.

Methodology

Dozens of BERT and other Transformer encoder architectures are available pre-trained from deep learning packages such as Tensorflow. Three distinct BERT models from TensorFlow Hub were selected, trained, and evaluated for effectiveness in classifying author vaccine status and side-effect reference, vaccine sentiment, and negative sentiment reasoning. Bert-cased uses the Bert-Base architecture from Devlin et al. (2019) with its preprocessing model preserving lower and upper case and accent markers before tokenization. BERT-uncased also uses Bert-Base but converts text to lower case and strips accent markers. Electra_base uses the ELECTRA-Base++ architecture from Clark et al. (2020) with the same preprocessing model as Bert_uncased.

Model training used a batch size of 16, with training performed over ten epochs using a learning rate of 3e-5 and AdamW as the optimizer. The architecture for each model consisted of a text input layer, a preprocessing layer, a ten percent dropout layer, and a dense layer of either Sigmoid or SoftMax activation, depending on whether the output label was binary or categorical. After training, the model weights with the lowest validation accuracies were reloaded and used in evaluation against the test dataset. In addition to recording accuracy and loss, micro and macro averages of model precision, recall, and F1 scores were recorded to assess performance.

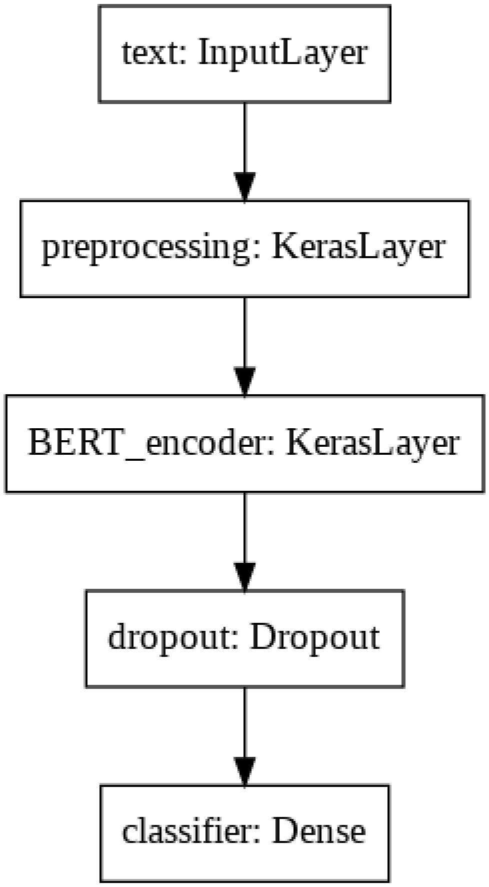

Figure 4. Pre-trained BERT classification model structure used for all tasks.

Results

Electra_base achieved the best performance for vaccination status classification with a test accuracy of 95%. Electra_base also produced the highest accuracy on side-effect reference classification, sentiment analysis, and negative sentiment reason classification, achieving accuracies of 89.8%, 79.8%, and 85.8%. Results indicate a fine-tuned Electra_base slightly outperforms other BERT models for COVID-19 Tweet classification tasks. A full table of evaluation results is available within Appendix C.

TensorHub Model Name	Fine-tuned Model Name	Task Name	Test Accuracy
bert_en_cased_L-12_H-768_A-12	bert_cased_m1	Vaccination Status Classification	91.0%
bert_en_cased_L-12_H-768_A-12	bert_cased_m2	Reference to Side-Effects Classification	87.3%
bert_en_cased_L-12_H-768_A-12	bert_cased_m3	Vaccination Sentiment Classification	77.0%
bert_en_cased_L-12_H-768_A-12	bert_cased_m4	Negative Sentiment Reason Classification	81.8%
electra_base	electra_m1	Vaccination Status Classification	95.0%
electra_base	electra_m2	Reference to Side-Effects Classification	89.8%
electra_base	electra_m3	Vaccination Sentiment Classification	79.8%
electra_base	electra_m4	Negative Sentiment Reason Classification	85.8%
bert_en_uncased_L-12_H-768_A-12	bert_uncased_m1	Vaccination Status Classification	92.3%
bert_en_uncased_L-12_H-768_A-12	bert_uncased_m2	Reference to Side-Effects Classification	87.5%
bert_en_uncased_L-12_H-768_A-12	bert_uncased_m3	Vaccination Sentiment Classification	77.8%
bert_en_uncased_L-12_H-768_A-12	bert_uncased_m4	Negative Sentiment Reason Classification	83.0%

Table 1. Summarized test dataset results for each model and annotation task.

Analysis

Task performance between each model was relatively similar, with Electra-base producing the best results. Each pre-trained model was able to achieve strong performance on high vaccination status and side-effect classification tasks. All three models were less effective in vaccination sentiment and negative sentiment reason classification tasks. However, every model achieved a respectable evaluation score when accounting for subjectivity and inconsistency in sentiment annotation. Using three classes of sentiment also presented complications for each model in determining sentiment. For Electra-base, only 6.5% of predicted negative or positive labels were labeled as the opposite sentiment.

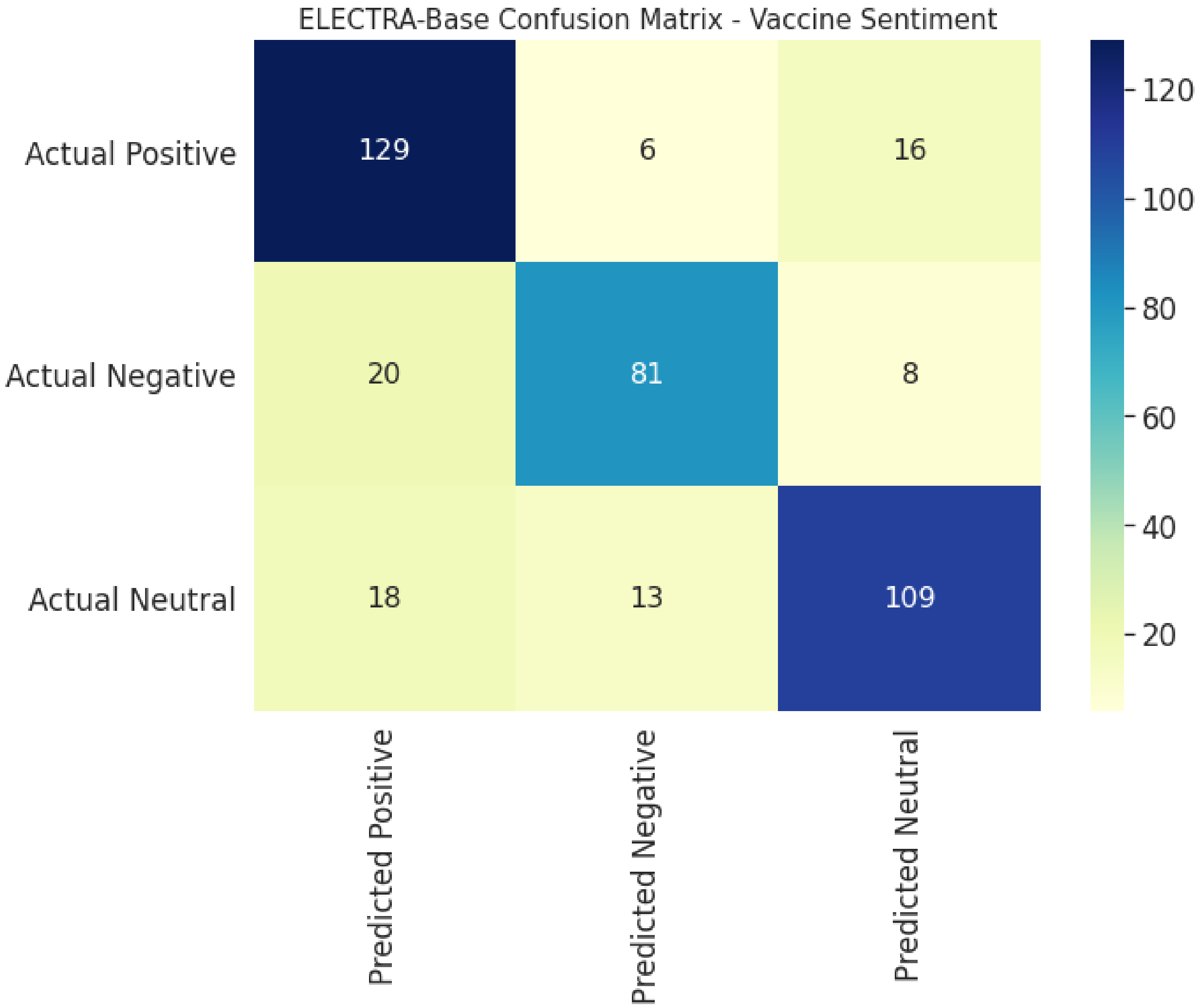

Figure 5. Confusion matrix visualizing Electra-base performance against test data.

On negative sentiment reason classification, each model achieved accuracy above 73%, which is attainable simply by labeling every tweet as 'Not Negative.' Accuracy above this level is significant considering the class imbalance and that a hierarchical model was not implemented to perform this task. Accuracy is likely to improve if negative sentiment reason is predicted only after completing sentiment classification to determine if tweets are positive, negative, or neutral.

Conclusions

Fine-tuned BERT models achieve high accuracy in classifying author vaccination status and reference to side-effects in Tweets related to COVID-19 vaccination. Models based on a pre-trained ELECTRA-Base++ architecture produced the highest accuracy. Results are respectable but less accurate for sentiment and negative sentiment reason classification tasks where there is a higher subjectivity to labeling. Experiments confirm BERT models provide an effective solution to the problem of processing millions of Tweets discussing COVID-19 vaccination for real-time monitoring and measurement of public sentiment and behavior. There is a significant opportunity to expand the scope of this work to improve performance or perform additional classification tasks on COVID-19 vaccination Tweets.

Directions for Future Work

Previous research on vaccine sentiment on social media benefited from the existence of existing annotated public corpora. With no corpora available for COVID-19 vaccination Tweets, the only solution was to build a new corpus through manual annotation. The corpus was not designed to determine whether tweets were related to COVID-19 vaccination as the standard API provided avenues to pre-filter tweets before annotation. However, the corpus could be expanded to include tweets unrelated to COVID-19 vaccination to train a fifth model to classify Tweets for relation or non-relation to COVID-19 vaccination.

Annotation is a time-intensive task, and a corpus of 2000 Tweets is a fraction of the volume of daily Tweets associated with COVID-19 vaccinations. While corpus size limitations were partially mitigated through the utilization of transfer learning using BERT, a larger corpus of annotated COVID-19 vaccine Tweets would improve the generalization of all classification models. Subjectivity and inconsistencies in the manual annotation of corpus Tweets are also worth addressing. Having multiple individuals review and annotate the same Tweets minimizes the risk that of Tweet mislabeling or instances where labels are applied inconsistently. Modifications to the annotation schema methodology, including refining criteria for positive, negative, or neutral sentiment annotation, may also address inconsistencies in Tweet labeling.

Experiments focused on fine-tuning a small set of pre-trained BERT models. The top-performing models on GLUE benchmarks provide the best catalog of models worth fine-tuning for COVID-19 vaccine Tweet classification tasks. Creating custom text pre-processing mechanisms is also an area of opportunity. Tweets' use of capitalization, emojis, and punctuation conveys essential information regarding topic sentiment, but most Tweets deviate from conventional grammar and syntax rules.

Incorporating COVID-19 vaccine annotation models into a complete tool or application represents the most significant area of opportunity for additional work. Using the Twitter API streaming services, Tweets related to COVID-19 vaccinations can be collected as Twitter users create them. As Tweets are collected, contents can be processed to automate annotation by using these classification models. Once the annotation is automated, data can be aggregated to monitor real-time changes in public vaccination sentiment. This data can serve as a mechanism to measure the effectiveness of vaccination campaign strategies and policies.

Code Referenced in this Document

Corpus data, as well as the code used for analysis, is provided in a repository at https://github.com/derekcarey/Covid-19_Vaccine_Tweet_Annotation_BERT . Inquiries should be directed to: derek.carey@outlook.com.

Appendix A: Distributions of COVID-19 Vaccine Tweet Corpus Datasets

Corpus Tweets were randomly shuffled before constructing training, validation, and test datasets. The training dataset used 1400 Tweets, with 200 Tweets used in the validation dataset and 400 Tweets used in the test dataset. Distributions of COVID-19 vaccine annotations for all datasets are provided in the figures below. For the vaccination_label, values of 0 mean the author's vaccine status is not vaccinated or unknown, and values of 1 represent the author is vaccinated. For sentiment_label, 0 means non-reference to vaccine side-effects, and 1 references vaccine side-effects. A label of 0 for sentiment_label indicates positive vaccine sentiment, with labels of 1 and 2 indicating negative and neutral sentiment. Negative_sentiment_reason_labels translates to 0 for safety, 1 for efficacy, 2 for availability, 3 for other, and 4 for "Not Negative."

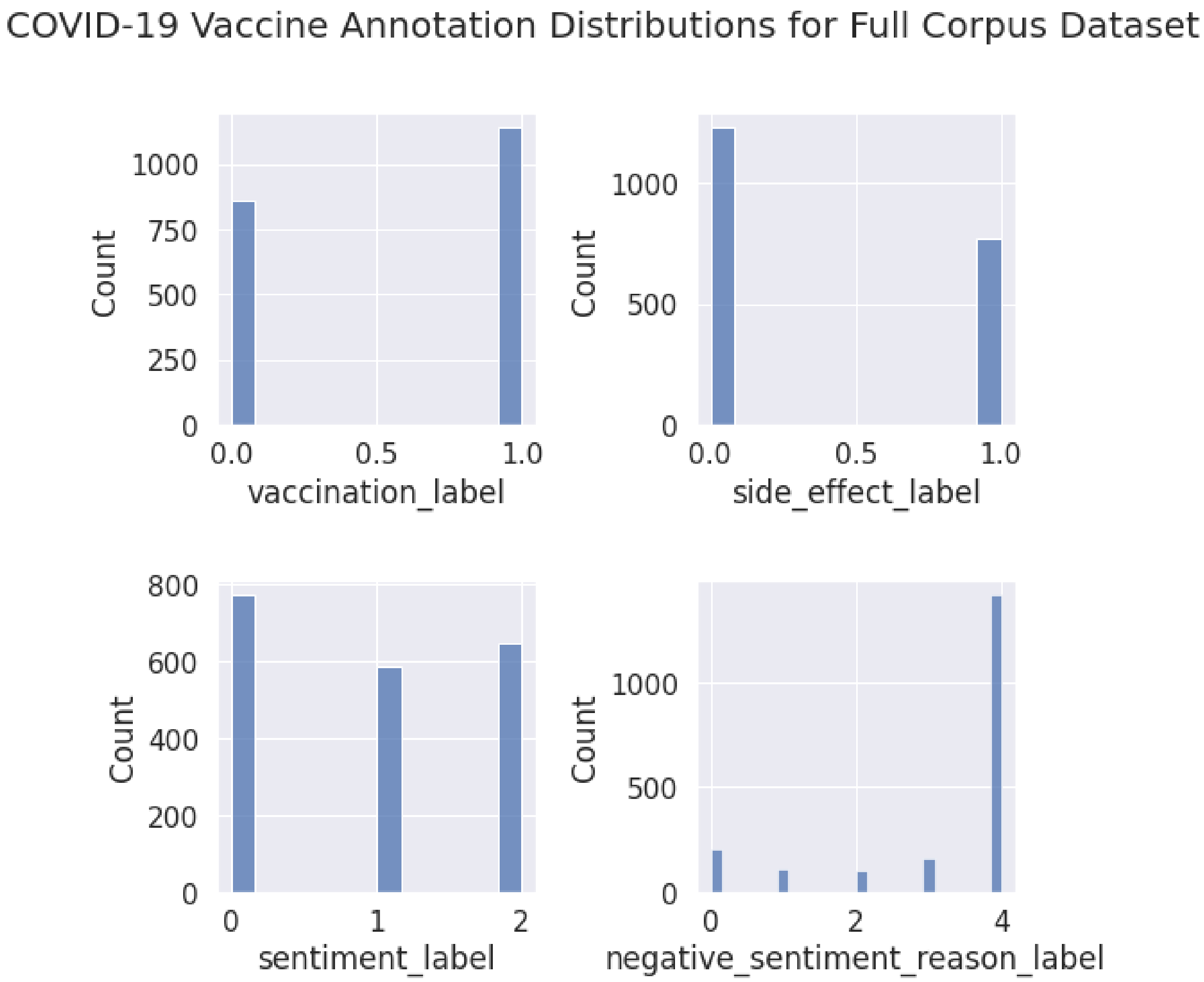

Figure 6. COVID-19 vaccine annotation distributions for the entire Tweet corpus.

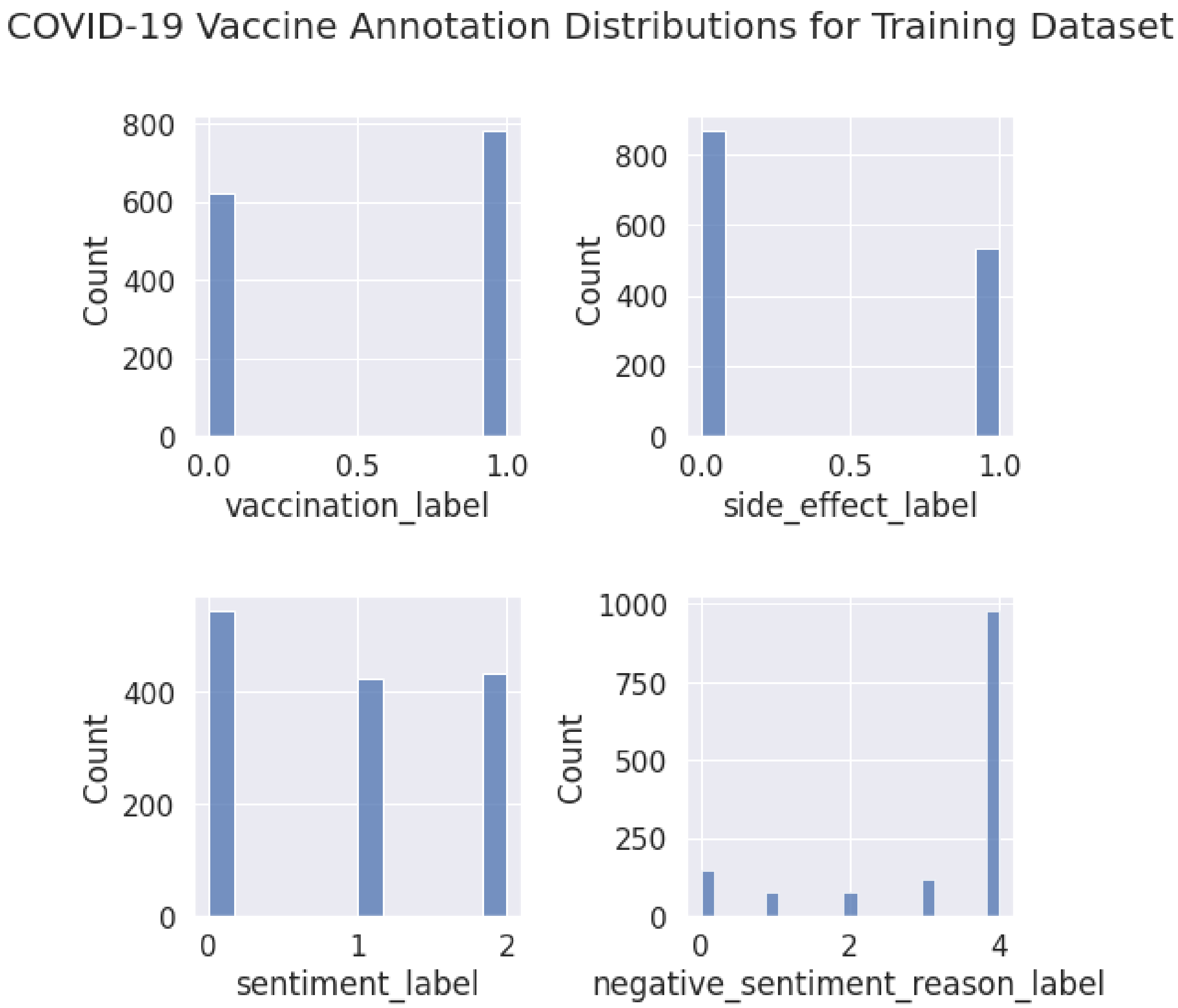

Figure 7. COVID-19 vaccine annotation distributions for the training dataset.

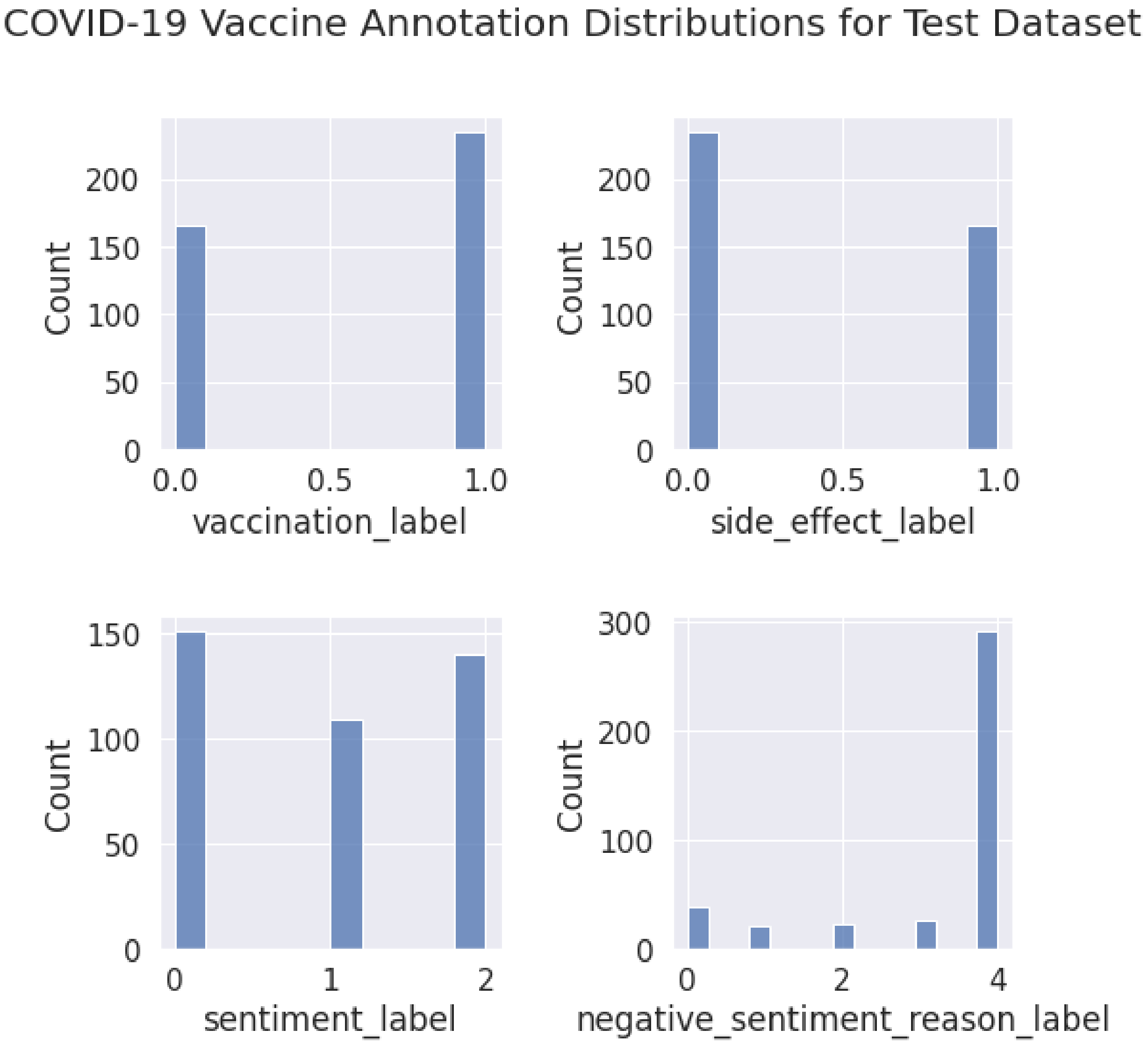

Figure 8. COVID-19 vaccine annotation distributions for the test dataset.

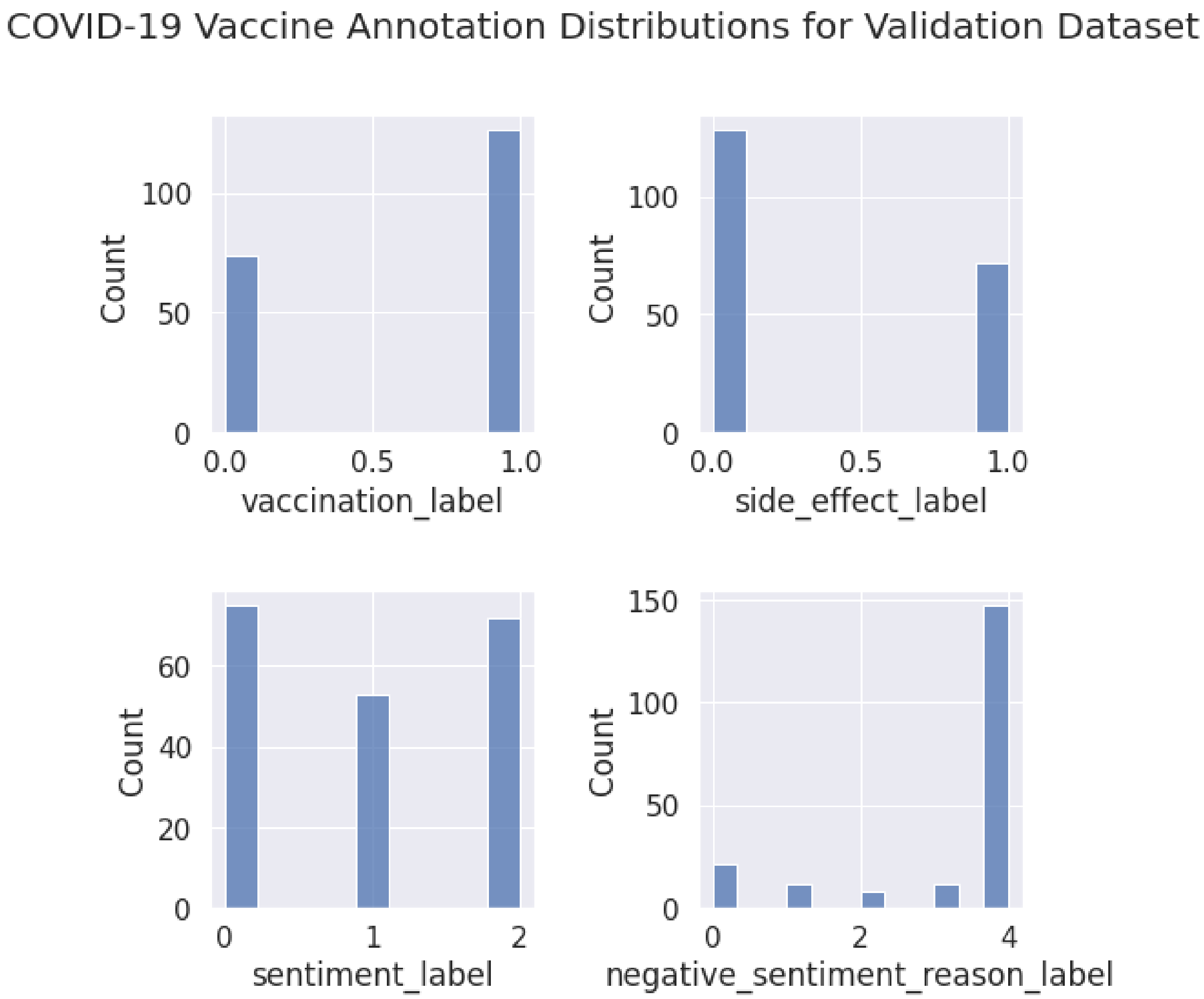

Figure 9. COVID-19 vaccine annotation distributions for the validation dataset.

Appendix B: Confusion Matrices for COVID-19 Vaccine Tweet Test Results

Confusion matrices were constructed to visually assess the performance of each of the three distinct BERT models across each of the four classification tasks.

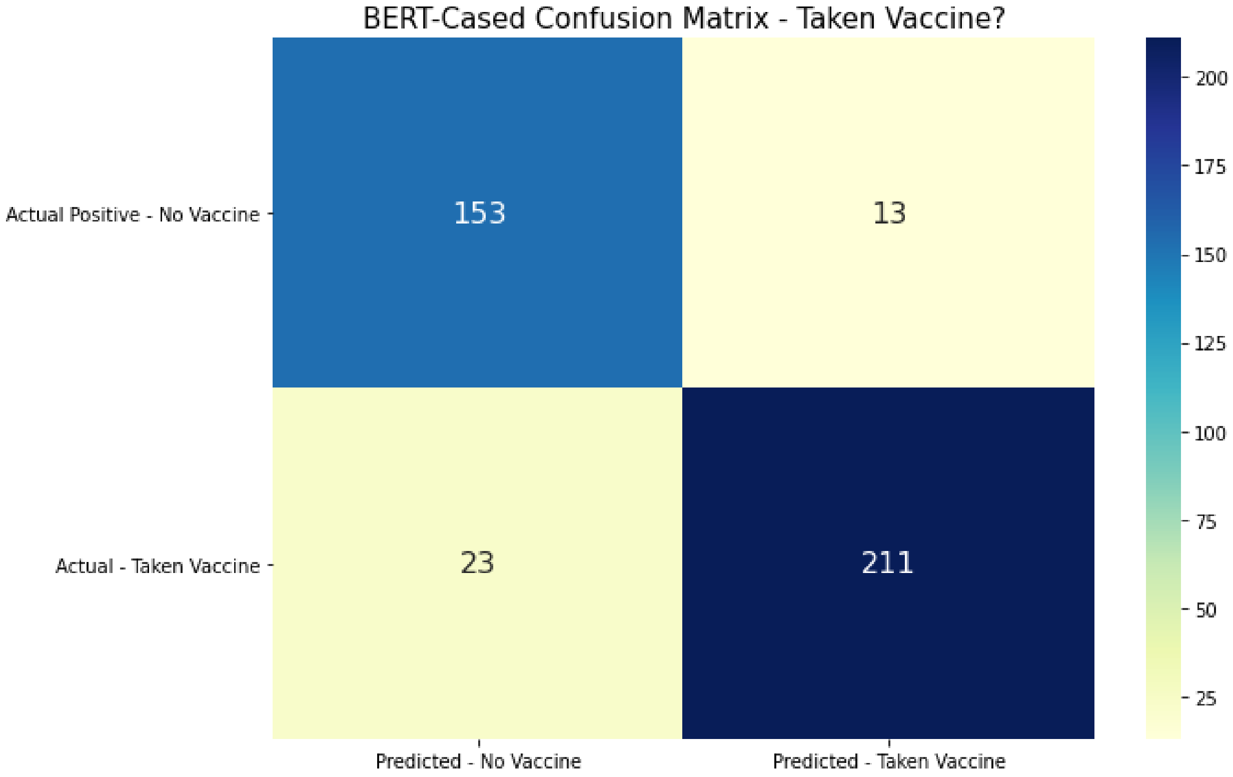

Figure 10. Test results for fine-tuned BERT-cased model on vaccination classification.

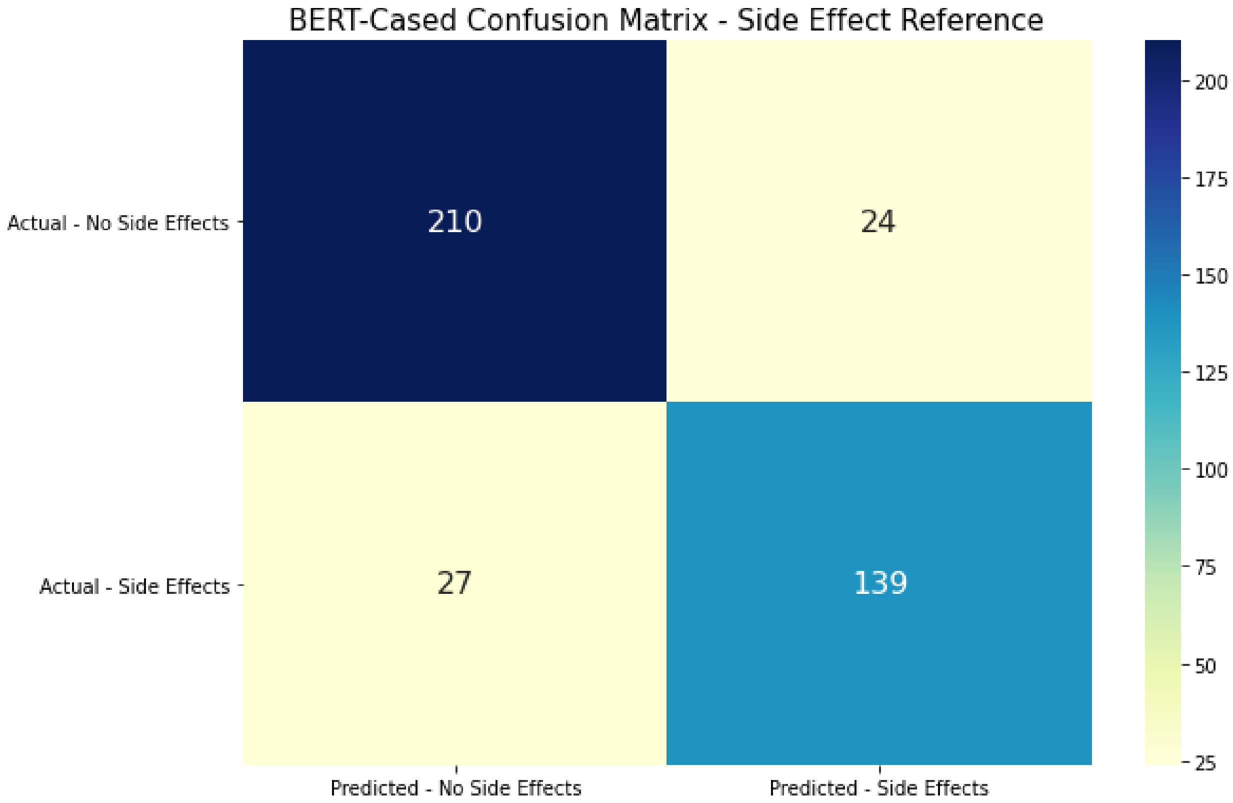

Figure 11. Test results for fine-tuned BERT-cased model on side-effect reference classification.

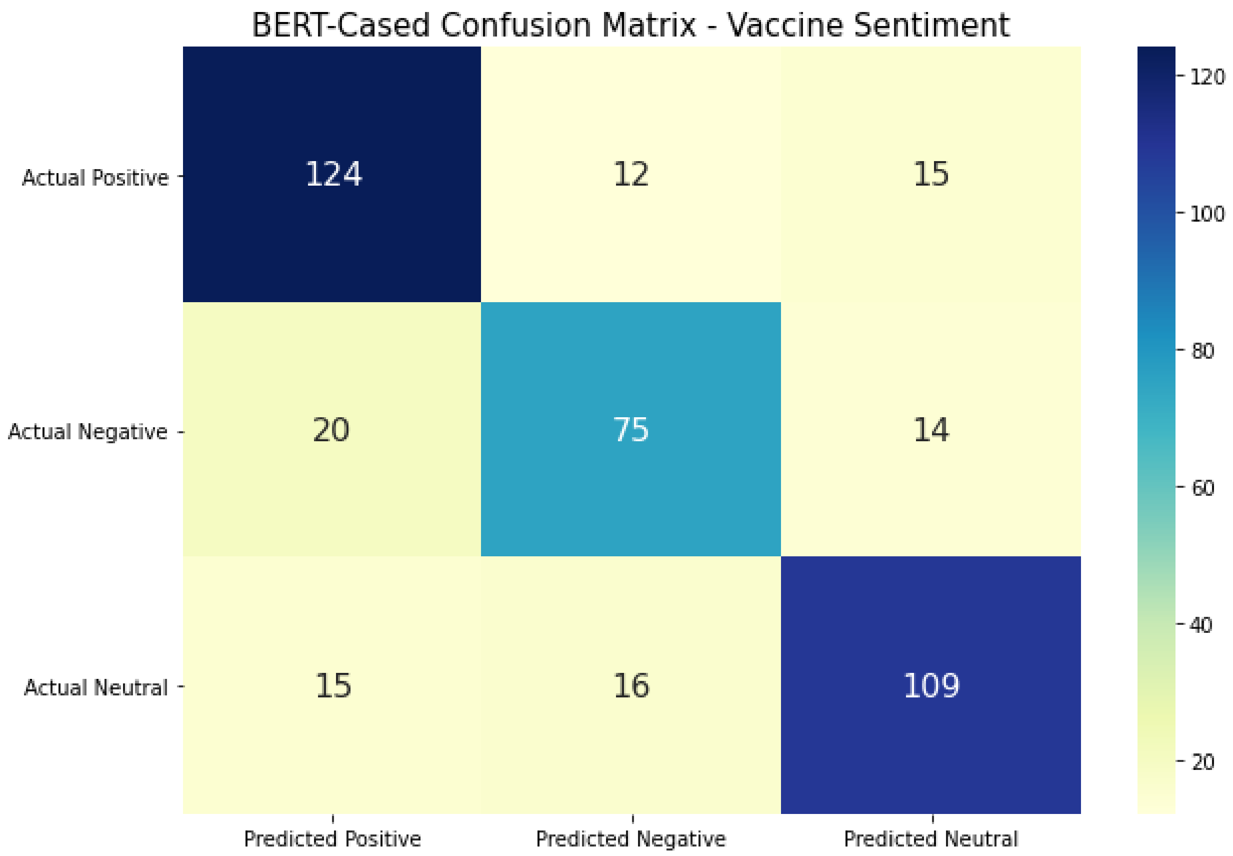

Figure 12. Test results for fine-tuned BERT-cased model on sentiment classification.

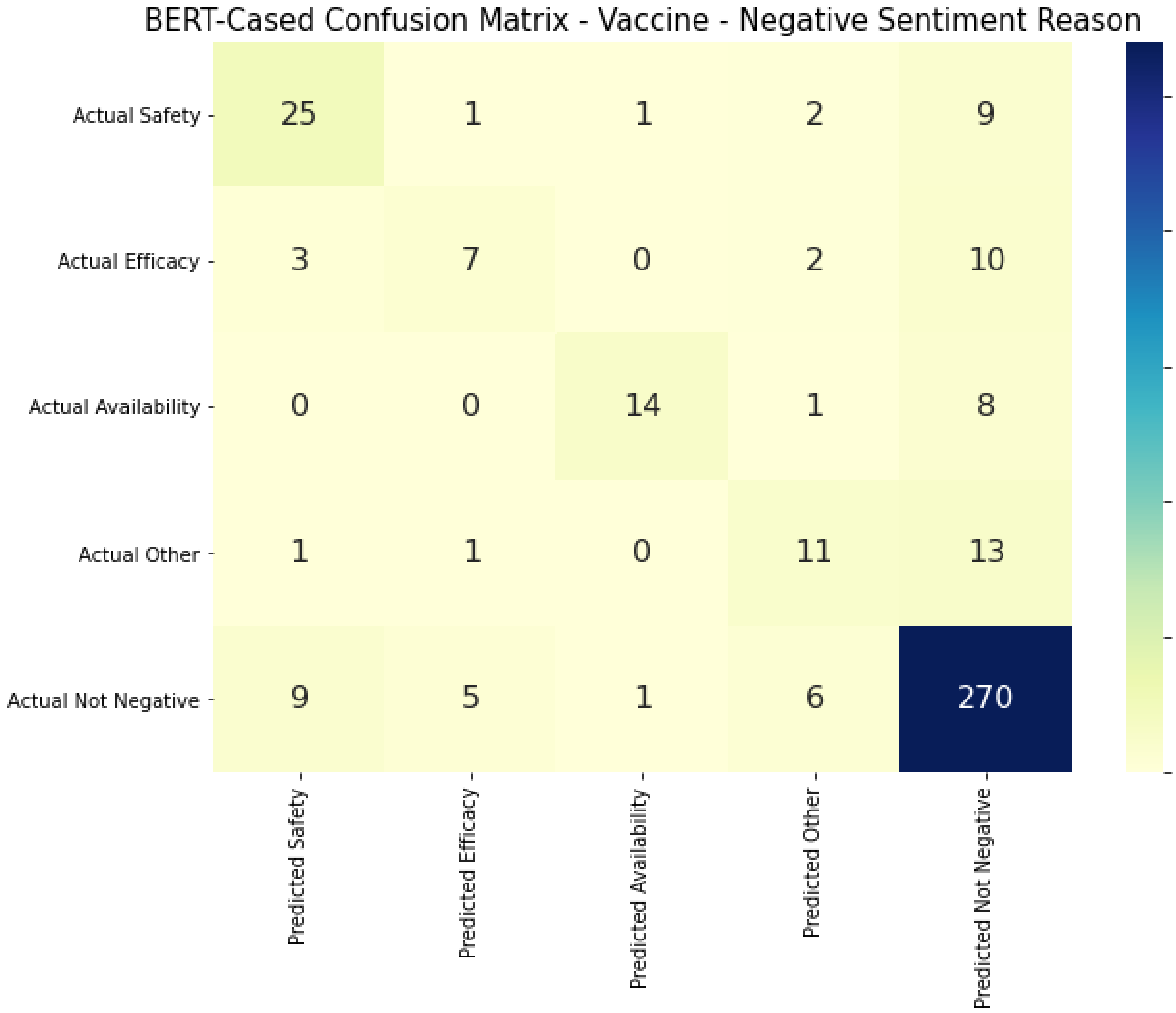

Figure 13. Test results for fine-tuned BERT-cased model on negative sentiment reason classification.

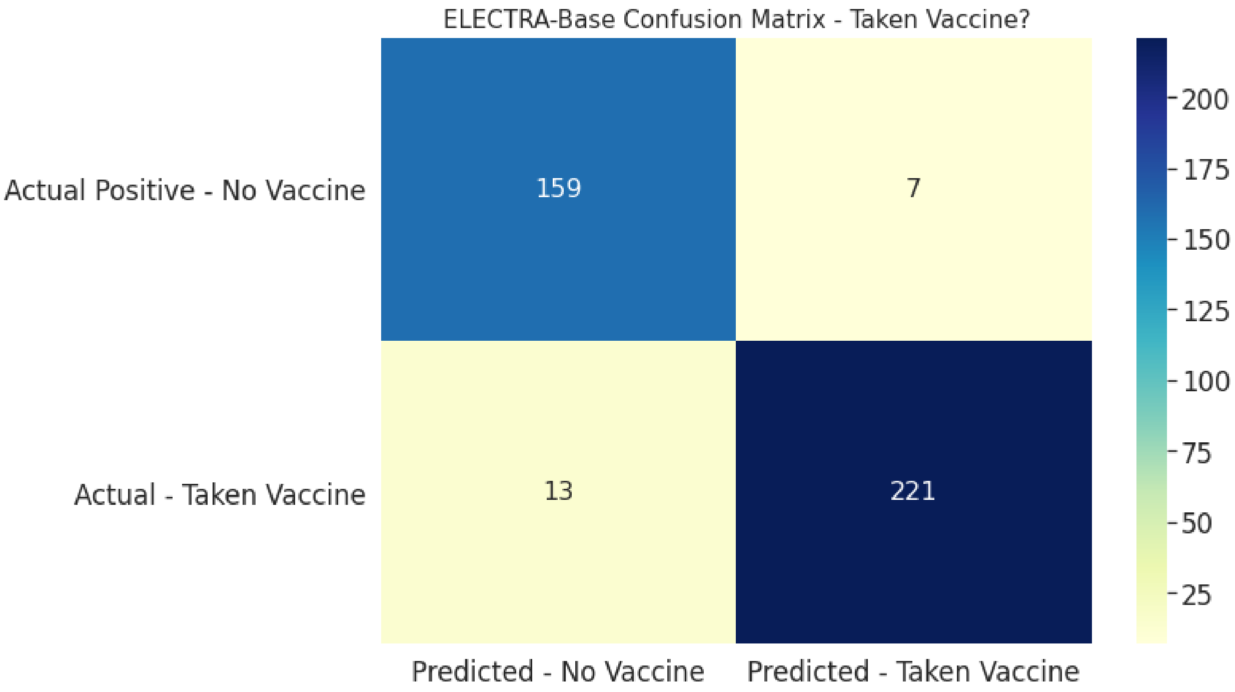

Figure 14. Test results for fine-tuned Electra-base model on vaccination classification.

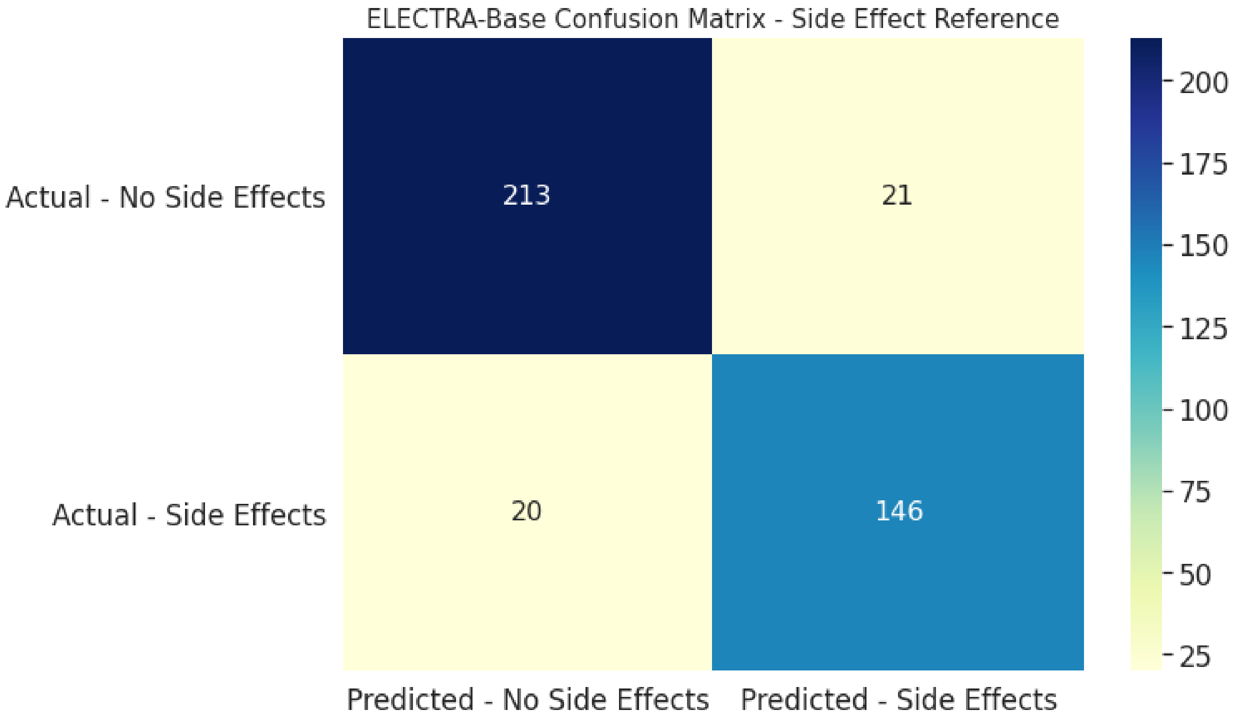

Figure 15. Test results for fine-tuned Electra-base model on side-effect reference classification.

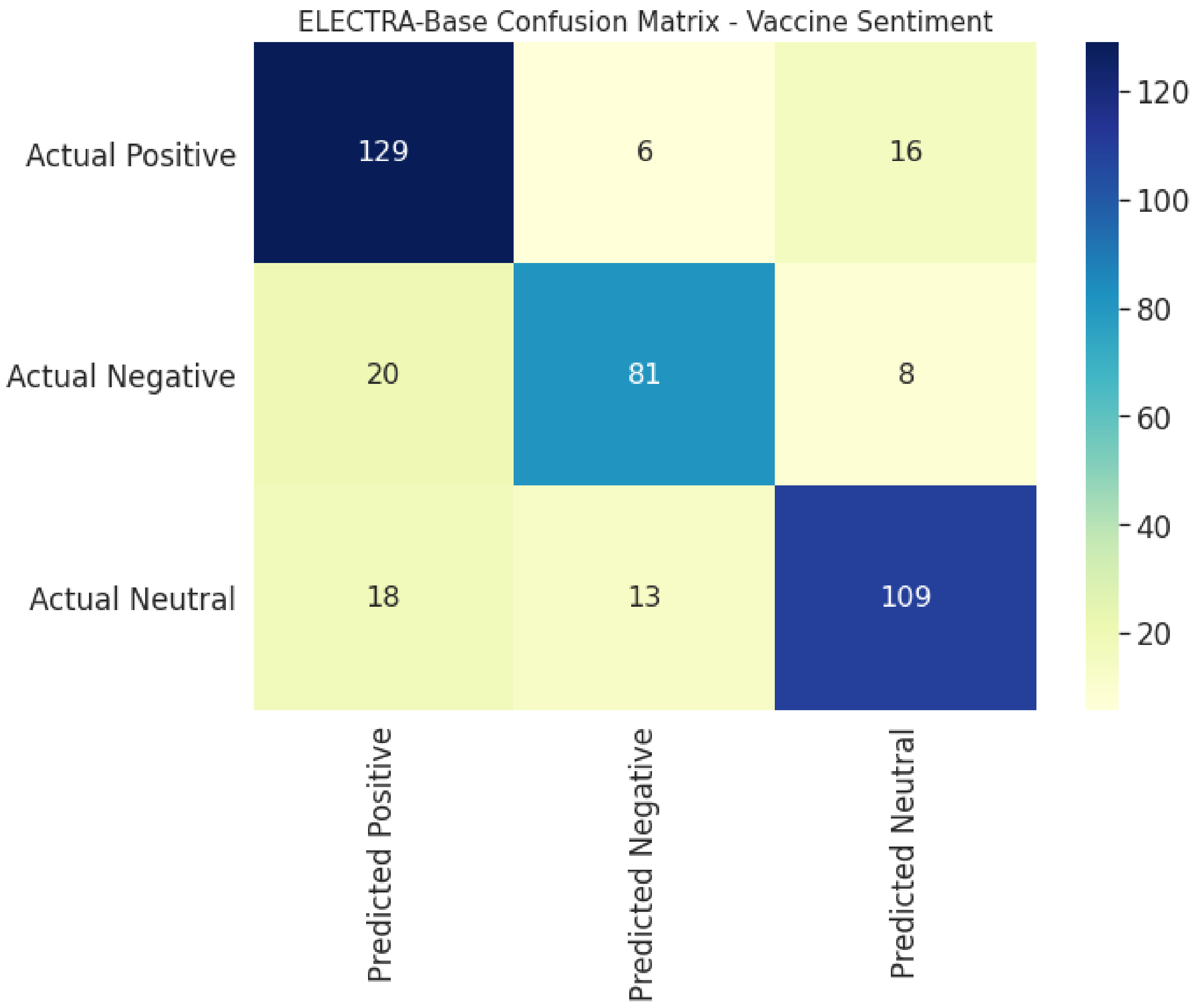

Figure 16. Test results for fine-tuned Electra-base model on sentiment classification.

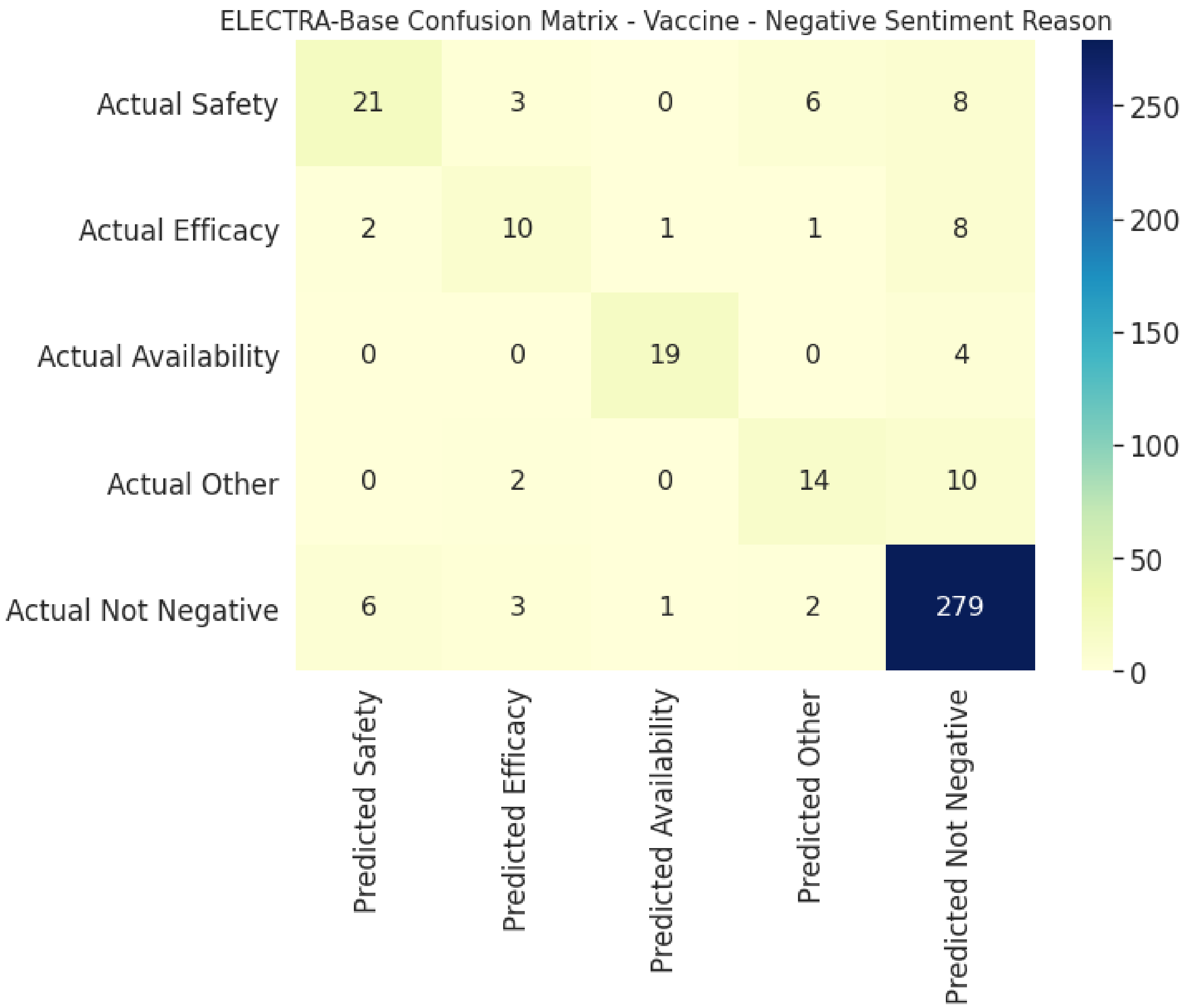

Figure 17. Test results for fine-tuned Electra-base model on negative sentiment reason classification.

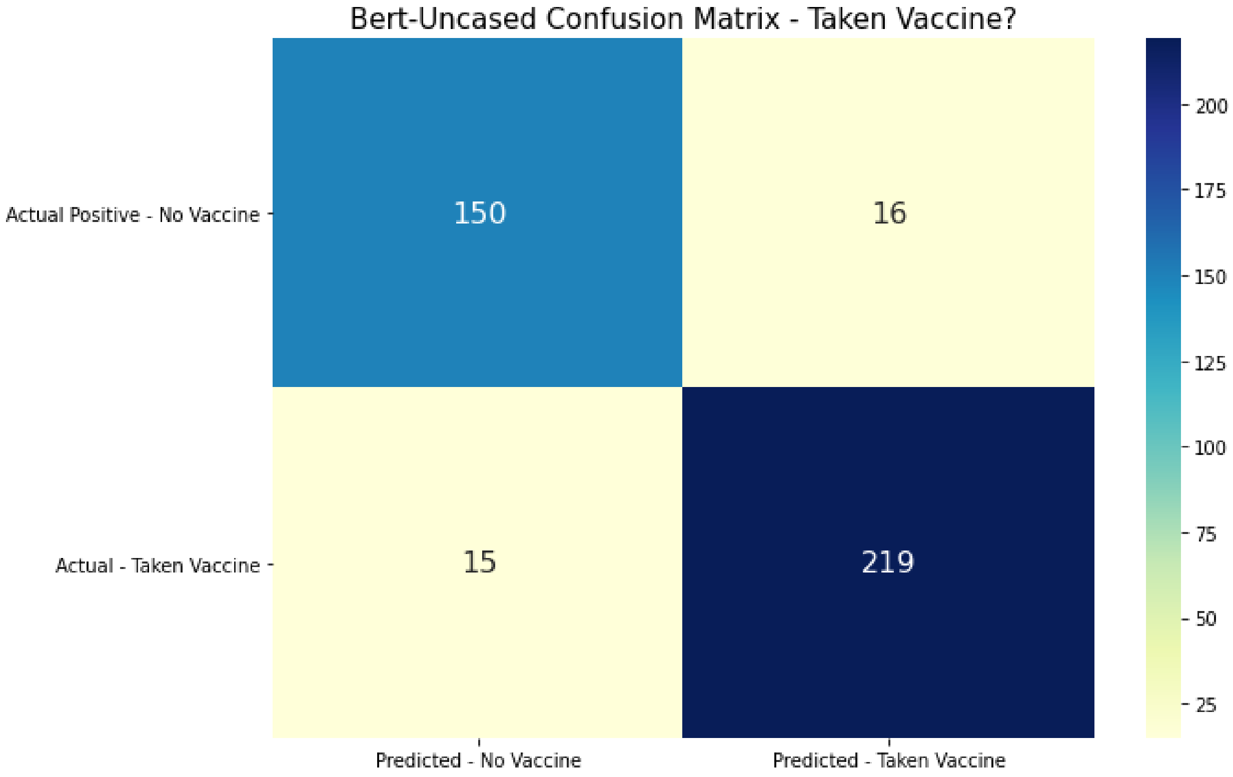

Figure 18. Test results for fine-tuned BERT-uncased model on vaccination classification.

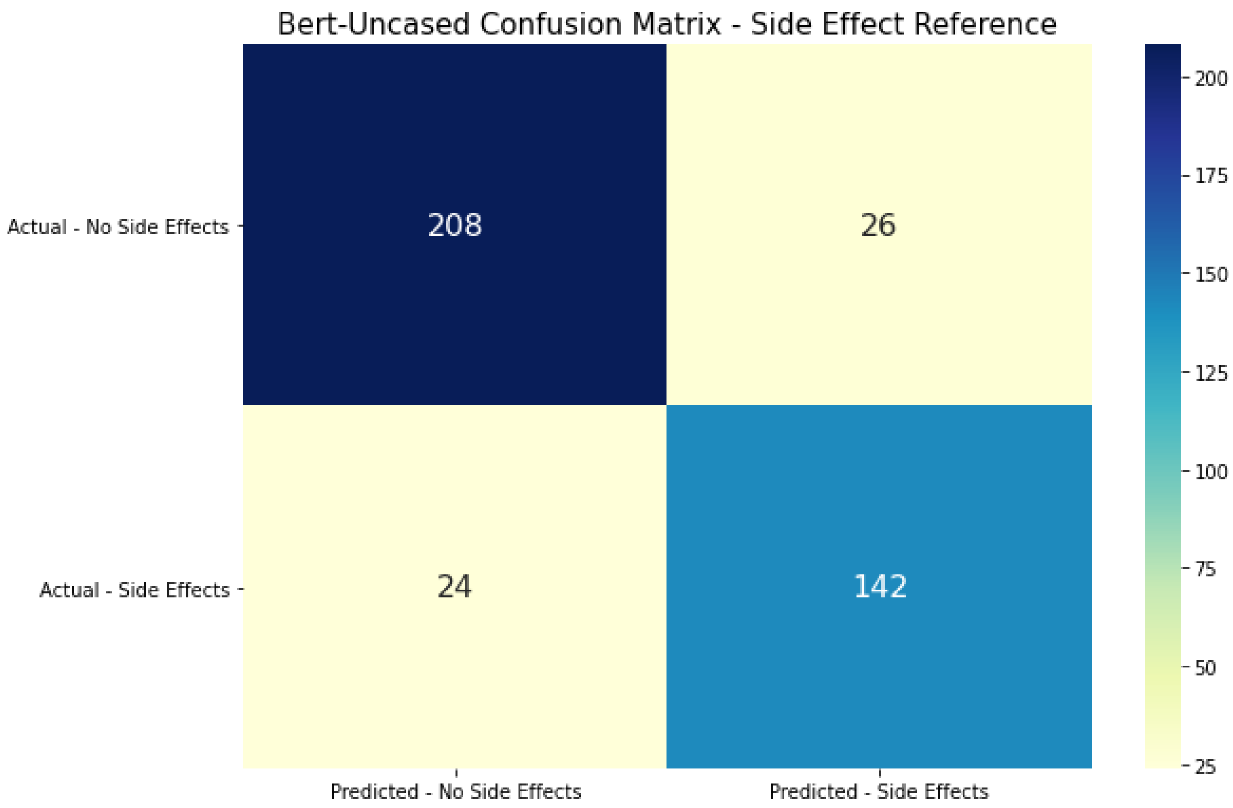

Figure 19. Test results for fine-tuned BERT-uncased model on side-effect reference classification.

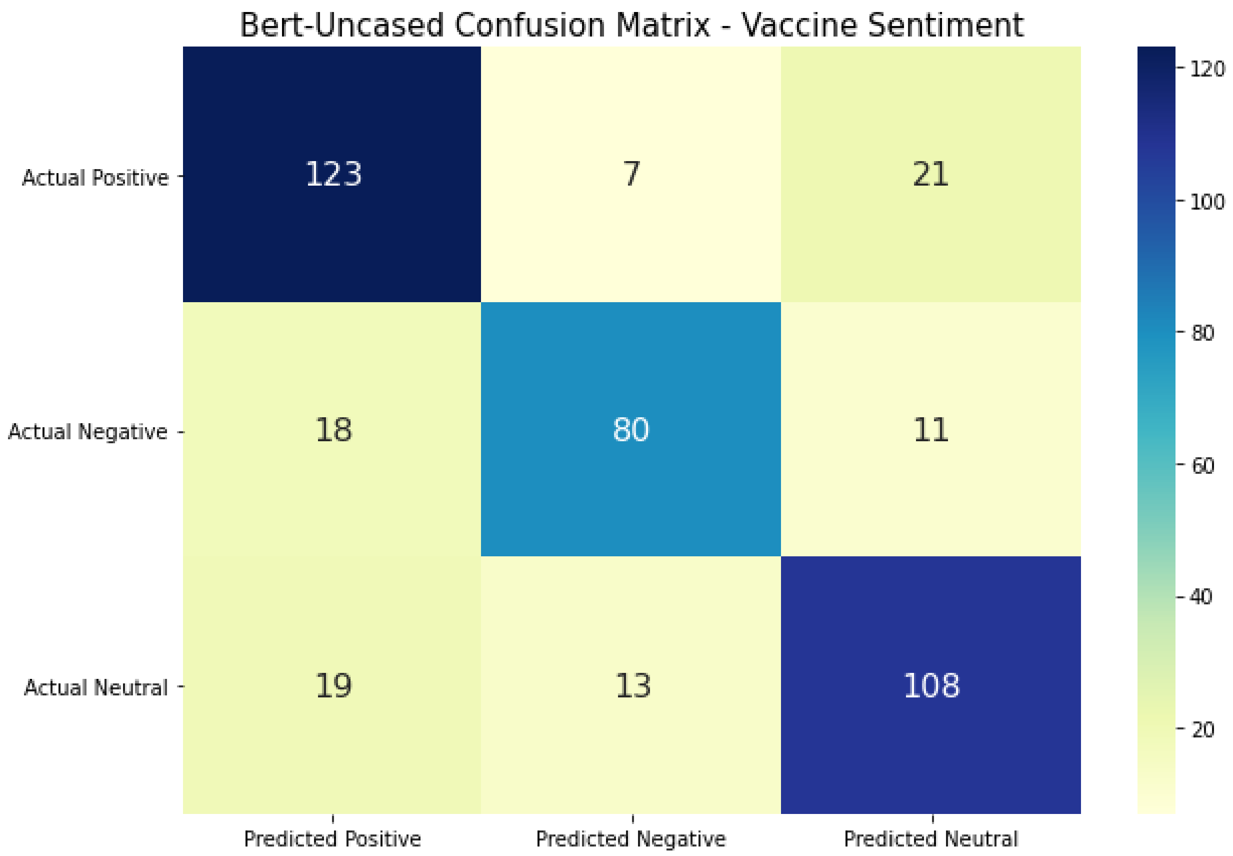

Figure 20. Test results for fine-tuned BERT-uncased model on sentiment classification.

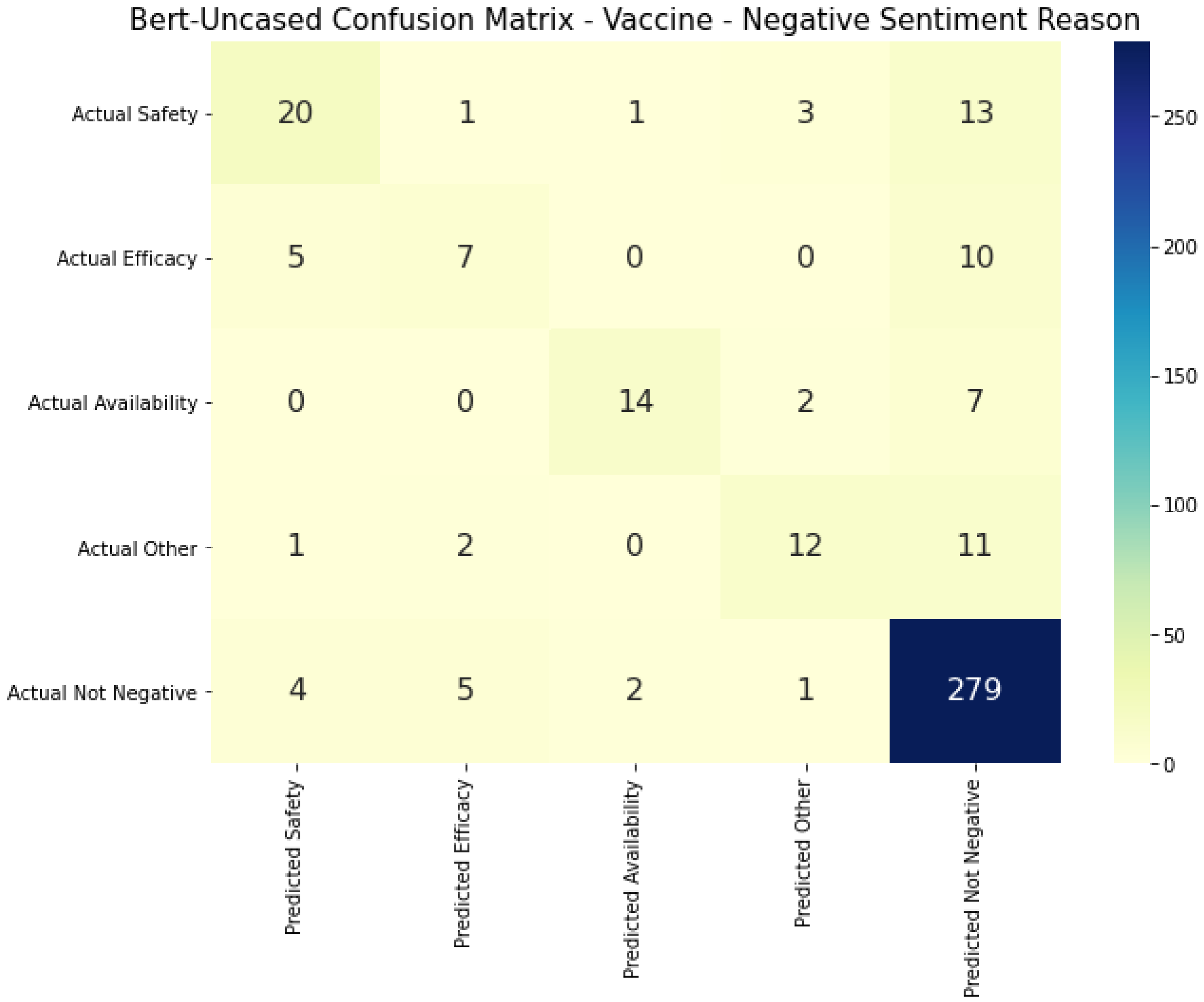

Figure 21. Test results for fine-tuned BERT-uncased model on negative sentiment reason classification.

Appendix C: Detailed Table of Model Evaluation Results

Several metrics were collected during test dataset evaluation of each pre-trained model against each of the four classification tasks. In addition to measuring accuracy, both micro and macro average precision, recall, and F1 scores were calculated. Macro-averages treat all classes equally and compute the metric independently for each category and then take the average. Micro-averages aggregate the contributions of all classes to compute the average and is useful for measuring results when there is class imbalance.

TensorHub Model Name	Classifier Task Name	Test Micro Precision	Test Micro Recall	Test Micro F1 Score	Test Macro Precision	Test Macro Recall	Test Macro F1 Score	Test Accuracy
bert_en_cased_L-12_H-768_A-12	bert_cased_m1	91.0%	91.0%	91.0%	90.6%	91.2%	90.8%	91.0%
bert_en_cased_L-12_H-768_A-12	bert_cased_m2	87.3%	87.3%	87.3%	86.9%	86.7%	86.8%	87.3%
bert_en_cased_L-12_H-768_A-12	bert_cased_m3	77.0%	77.0%	77.0%	76.6%	76.3%	76.4%	77.0%
bert_en_cased_L-12_H-768_A-12	bert_cased_m4	81.8%	81.8%	81.8%	68.1%	58.7%	62.4%	81.8%
electra_base	electra_base_m1	95.0%	95.0%	95.0%	94.7%	95.1%	94.9%	95.0%
electra_base	electra_base_m2	89.8%	89.8%	89.8%	89.4%	89.5%	89.5%	89.8%
electra_base	electra_base_m3	79.8%	79.8%	79.8%	80.1%	79.2%	79.5%	79.8%
electra_base	electra_base_m4	85.8%	85.8%	85.8%	73.9%	66.6%	69.8%	85.8%
bert_en_uncased_L-12_H-768_A-12	bert_uncased_m1	92.3%	92.3%	92.3%	92.1%	92.0%	92.0%	92.3%
bert_en_uncased_L-12_H-768_A-12	bert_uncased_m2	87.5%	87.5%	87.5%	87.1%	87.2%	87.2%	87.5%
bert_en_uncased_L-12_H-768_A-12	bert_uncased_m3	77.8%	77.8%	77.8%	78.0%	77.3%	77.6%	77.8%
bert_en_uncased_L-12_H-768_A-12	bert_uncased_m4	83.0%	83.0%	83.0%	69.9%	57.5%	62.5%	83.0%

Table 2. Full test dataset results for each model and annotation task including micro and macro average precision, recall, and F1 scores.

References

"The COVID-19 US Vaccine Sentiment Series." *United States - EN*. https://www.bcg.com/en-us/publications/2021/covid-19-us-vaccine-sentiment-series.

Blume, Stuart S., Christine Holmberg, and Paul Greenough. 2017. *The Politics of Vaccination: a Global History*. Manchester University Press.

Clark, Kevin, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. "ELECTRA: Pre-Training Text Encoders as Discriminators Rather Than Generators." *ArXiv.org*. March 23. https://arxiv.org/abs/2003.10555.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee and Kristina Toutanova. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." *ArXiv.org*, May 24, 2019. https://arxiv.org/abs/1810.04805.

Du, Jingcheng, Jun Xu, Hsingyi Song, Xiangyu Liu and Cui Tao. "Optimization on Machine Learning Based Approaches for Sentiment Analysis on HPV Vaccines Related Tweets." *Journal of Biomedical Semantics* 8, no 1 (2017). doi:10.1186/s13326-017-0120-6.

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer and Veselin Stoyanov. "RoBERTa: A Robustly Optimized BERT Pretraining Approach." *ArXiv.org*, July 26, 2019. https://arxiv.org/abs/1907.11692.

Piedrahita-Valdés, Hilary, Diego Piedrahita-Castillo, Javier Bermejo-Higuera, Patricia Guillem-Saiz, Juan Ramón Bermejo-Higuera, Javier Guillem-Saiz, Juan Antonio Sicilia-Montalvo and Francisco Machío-Regidor. "Vaccine Hesitancy on Social Media: Sentiment Analysis from June 2011 to April 2019." *Vaccines* 9, no 1 (2021), 28. doi:10.3390/vaccines9010028.

"Ten Health Issues WHO Will Tackle This Year." *World Health Organization*. World Health Organization. https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019.

"TensorFlow Hub." *TensorFlow*. https://www.tensorflow.org/hub.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser and Illia Polosukhin. "Attention Is All You Need." *ArXiv.org*, December 6, 2017. https://arxiv.org/abs/1706.03762.

Wang, Alex, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy and Samuel Bowman. "GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language

Understanding." *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, 2018. doi:10.18653/v1/w18-5446.

Zhang, Li, Haimeng Fan, Chengxia Peng, Guozheng Rao and Qing Cong. "Sentiment Analysis Methods for HPV Vaccines Related Tweets Based on Transfer Learning." *Healthcare* 8, no 3 (2020), 307. doi:10.3390/healthcare8030307.

Zhu, Yukun, Ryan Kiros, Rich Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba and Sanja Fidler. "Aligning Books and Movies: Towards Story-Like Visual Explanations by Watching Movies and Reading Books." *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015. doi:10.1109/iccv.2015.11.

Zimmer, Carl, Jonathan Corum and Sui-lee Wee. "Coronavirus Vaccine Tracker." *The New York Times*. The New York Times, June 10, 2020. https://www.nytimes.com/interactive/2020/science/coronavirus-vaccine-tracker.html.