Mining Epilepsy Diagnosis Information from EMU Notes, with NLP

Justina Lam, SEAS 2025
Dr. Colin Ellis, M.D., Kevin Xie, Dr. Erin Conrad, M.D., Nina Ghosn
Center for Neuroengineering and Therapeutics, Perelman School of Medicine, School of Engineering and Applied Science
Penn Undergraduate Research Mentoring Program (PURM)

Introduction and Objectives

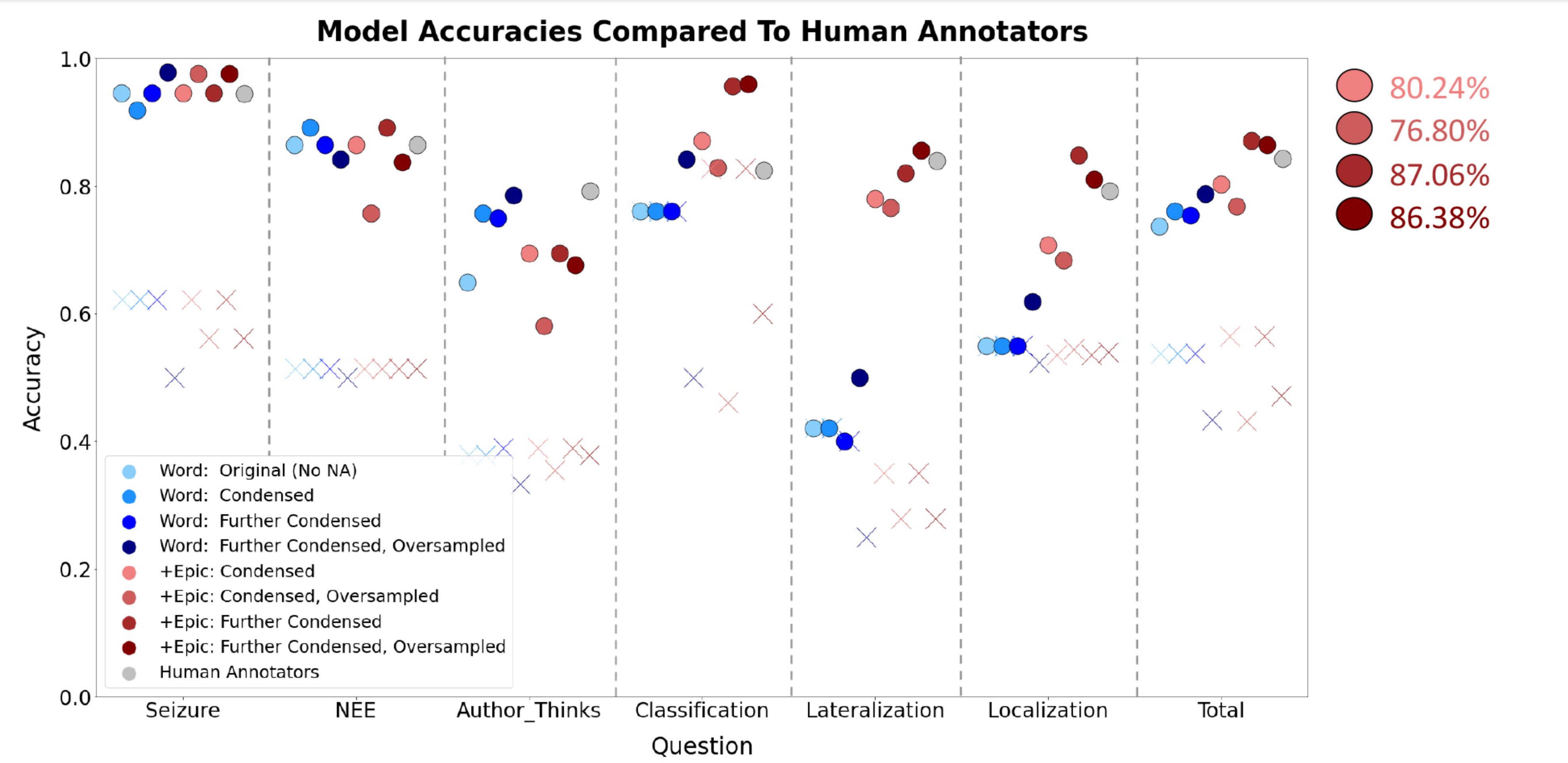
- Natural Language Processing (NLP) is a type of machine learning widely used for tasks involving text, speech, and other forms of human communication.
- Human writing and other forms of natural language are extremely diverse in style, wording, sentence structure, and other characteristics, as well as semantic ambiguity.
- In gathering information from this type of irregular data, NLP models learn to interpret this media and recognize what features are important, proving much more successful than rules-based approaches that do not use machine learning.
- In this project, we trained an NLP model to analyze patient notes from Epilepsy Monitoring Unit (EMU) admissions and extract information relating to each patient's diagnosis, as well as several characteristics of their epilepsy, if applicable.
- Our objective was to train a classification model to answer the following six classification questions about each patient note:

1) Did the patient have an epileptic seizure during this visit?
2) Did the patient have non-epileptic events during this visit?
3) Did the author conclude that the patient may have epilepsy?

And if applicable:

4) What is the classification of the patient's epilepsy?
5) What is the lateralization of the patient's epileptic seizures?
6) What is the localization of the patient's epileptic seizures?

Results: Accuracies of Classification



- The best-performing model was the one trained on the larger dataset that had undergone two rounds of label condensing but no oversampling; this had a slightly higher total accuracy (87.06%) than the one trained on a dataset that was oversampled as well as condensed twice (86.38%), likely due to overfitting.
- The expansion of the training dataset had an extremely significant impact on accuracy, especially for the last three questions (which had more possible answers to choose from). Oversampling the less common categories increased accuracy for the smaller initial training dataset, but had a less dramatic effect on the larger one. (X symbols represent the accuracy of a model that simply guesses the most common label for each question.)

Applied Model: Prediction Distributions

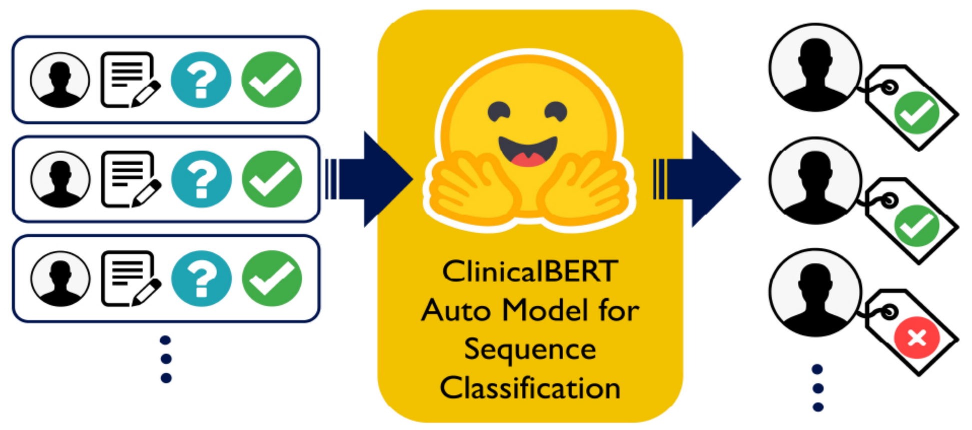
1035 Epilepsy Unlikely	52 Possible Epilepsy	1534 Certain/Likely Epilepsy
920 Had NEE	1216 Had Epileptic Seizures	
2621 Total Files Predicted		

448 Right Hemisphere	31 Both	795 Left Hemisphere
872 Temporal Lobe Onset	1749 Other/Unknown Localization	
2210 Predicted to Have Focal Epilepsy		

- Having determined the overall best-performing model, we ran this model on a dataset of over 2600 other patient notes, and determined the distribution of predicted answers to each question.
- The distribution of predictions was consistent with that of the randomly sampled training dataset for most questions. The model that performed second best overall (trained on data both condensed and oversampled) produced a similar prediction distribution, with higher counts of labels for "Possible" epilepsy and "Both" left and right seizure lateralization for questions 3 and 5, respectively, as well as fewer "Focal" predictions for question 4—the expected results of oversampling less common categories.

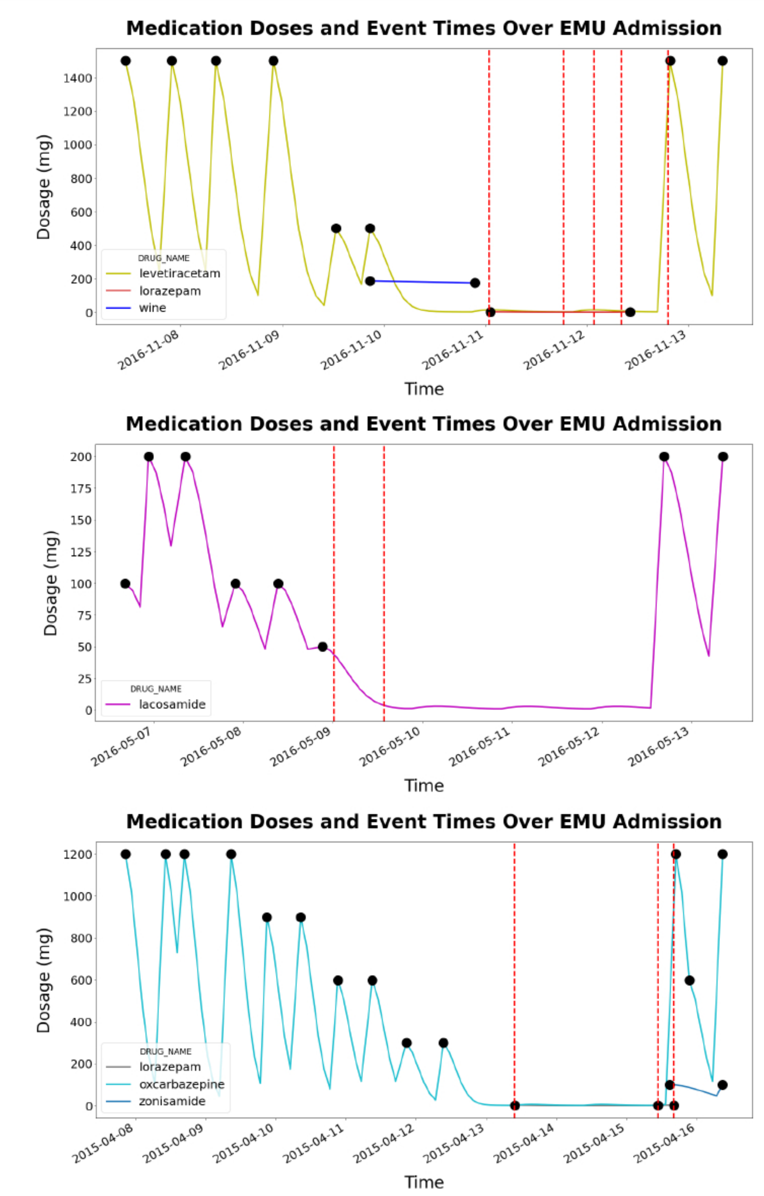
Methods and Data Augmentation

- From a database of over 2500 patient notes from the EMU, we randomly selected 124 notes to manually annotate as training data for the model, and later added 545 additional labeled training notes from the Epic platform.
- For each of these, we manually answered the six questions above, regarding the patient's diagnosis. For each question, we gave each note a label corresponding to the "correct" ground-truth adjudicated answer according to human annotators.
- We then provided this labeled training data to Hugging Face's Auto Model for Sequence Classification, which uses the ClinicalBERT transformer model pre-trained on medical notes.
- In order to improve the model's accuracy, we made several modifications to the training dataset:
  1. Condensing less common categories, or answer labels, to create a more balanced training dataset and reduce bias toward the more common classes
  2. Oversampling to further equalize the amount of training data with each label, within each question



Extracted Seizure Times and Medications

- In addition to training the model to extract information about each patient's diagnosis from each EMU note, we also extracted the precise dates and times of their seizures.
- We then used a pharmacokinetic model for the absorption and metabolism of anti-seizure medications, developed by another researcher at the lab, to calculate the blood concentration over time of each medication administered during the visit.
- By plotting these alongside seizure times over the course of each patient admission, we can visualize the concentration of each medication when the first seizure occurred.
- This allows us to evaluate the effectiveness of seizure induction procedures such as medication tapering and alcohol administration, and optimize them for greater likelihood of successful seizure induction.



Conclusions and Next Steps

- Increasing the size of the training dataset, and condensing less common categories to create a balanced training set, had the most significant impact on model accuracy.
- Further improvements to prediction accuracy may include exploring different methods of data augmentation, rather than oversampling and simple duplication of notes with less common labels, as well as adding additional training data.
- The plots of anti-seizure medication (ASM) dosages over time have also been modified to calculate and graph the blood concentration (mg/L) of each medication over time, rather than dosage (mg). Future research may explore correlations between the blood concentration of each ASM at the time of first seizure, medication tapering speed, administration of alcohol as a seizure induction method, and other factors, with whether seizures were successfully induced, in order to optimize the seizure induction process and treatment of EMU patients.

Acknowledgements and References

- Xie, Kevin, et al. "Extracting Seizure Frequency from Epilepsy Clinic Notes: A Machine Reading Approach to Natural Language Processing." *Journal of the American Medical Informatics Association*, vol. 29, no. 5, 2022, pp. 873–881., https://doi.org/10.1093/jamia/ocac018.
- I would like to thank the researchers at the Center for Neuroengineering and Therapeutics for their guidance and mentorship this summer, as well as the Penn Undergraduate Research Mentoring Program (PURM).