A machine learning-based approach to classifying a provider's description of chest pain

Elie Donath MD MPH[1], Yao Li MD[1], Nick Cheney PhD[1]
[1] University of Vermont

Background

Five million patients present to an emergency department in the US annually with a chief complaint of chest pain. Attempts to classify chest pain as either "cardiac" vs "noncardiac" or "typical" vs "atypical" are based on standards set decades ago and may be outdated. Despite risk stratification tools for patients with intermediate or high risk for major cardiac events, 2-3% of these patients will be discharged home and subsequently diagnosed with acute coronary syndrome (ACS) within one week. Delayed diagnoses are a major risk factor for heart failure, arrhythmias, and death. The potential for these adverse outcomes must be weighed against the 70-80% of these patients found to have a non-emergent, and often noncardiac, cause for their symptoms. Little is known about the extent by which machine learning (ML)-based approaches involving text data may help better delineate these patient populations. Electronic medical records contain reams of structured data including numerical (e.g. vitals, lab values) or categorical (e.g., gender, comorbidity) variables that are simple to integrate into algorithmic models. In contrast, there is no established consensus on how to effectively process unstructured data (e.g., free text notes). As a result – despite their considerable potential value - very little free text is being processed, interrogated, or harnessed in any meaningful way. These notes are rich in clinical and demographic data; they describe a patient's current problems, treatment history, progress, and rationale for clinical decisions. The past several years has seen an explosion in natural language processing (NLP)-based techniques enabling highly sophisticated analytical approaches to ingesting text data. As gaps still exist in our understanding of the different ways that cardiac vs noncardiac chest pain can manifest, we hope to ascertain whether these advances in NLP will help differentiate between cardiac and non-cardiac chest pain presentations, augment the medical decision-making process, and improve existing risk stratification methods in a cost-effective manner.

Methods

Data is obtained from the Medical Information Mart for Intensive Care III (MIMIC-III) database. MIMIC-III includes over 58,000 hospital admissions from over 40,000 patients across an eleven-year period at Beth Israel Deaconess Medical Center. There are 2,083,180 de-identified notes associated with these admissions. Of these, 141,624 were physician notes and 6,600 of those contained the words "chest pressure" or "chest pain". 2,877 of these notes included the use of "chest pain" or "chest pressure" as a pertinent negative and these were subsequently removed. Of the 3,723 notes that remained, 2,732 were associated with a first-time troponin value obtained within 12 hours of that note. We considered a troponin above 0.1ng/ml (adjusted for renal function) to be significant and indicative of ACS for the purpose of this research. After removing notes that contained duplicated descriptions of chest pain (notes are sometimes copied-and-pasted throughout charts) and cases in which multiple notes are written during the same time frame and hospitalization (we only use the first chronological descriptor), we were left with 837 unique descriptions of chest pain (that had an associated troponin value obtained within 12 hours). In cases where multiple troponin values were obtained, we filtered for the maximum troponin value. Standard preprocessing techniques (punctuation/digit removal, lowercasing text, etc…) were then deployed. A variety of text-classifying algorithmic schemes were implemented including more traditional ML approaches (like a term frequency-inverse document frequency/bag of words model followed by Naïve Bayes/Logistic Regression/SVC) as well as more novel transformer-based techniques including bidirectional encoder representations from transformers (BERT) and ClinicalBERT (which is pretrained on clinical text as opposed to general text). For validation and testing, the cohort was split into five folds (in each fold 10% is used for validation, 10% for the test set and the other 80% for training) and the test set from each of the models was evaluated according to precision, recall and F1-Score.

Results

Of 837 instances of chest pain, 273 (32.6%) were associated with a clinically significant elevated troponin, which we considered – for the purpose of this research – to represent true "cardiac chest pain". The average troponin value, across all 837 instances, was 0.87ng/ml (SD 2.08, range 0.0-19.1). Using a more conventional ML-based approach, we achieved a wide range of F1 scores (from 63% to 94% depending on ML model) on the test set. With novel transformers-based models, we achieved consistently higher performance (ranging from 89% to 94% depending on model). With traditional ML models, we were able to identify unigrams and bigrams associated with highest/lowest word coefficients. We found that words like "nstemi", "plavix" and "heparin gtt" were closely associated with "cardiac chest pain" whereas words like "egd", "anemia", "transfuse" and "ct" were more closely associated with "noncardiac chest pain".

Conclusion

This research hints at the tremendous value associated with deploying ML/DL-based approaches towards interrogating unstructured clinical text data. We anticipate deploying similar approaches towards larger datasets and more fine-grained set of clinical classification schema.