Using Simulated Patient-Doctor Conversations for Respiratory Disease Classification

Ricardo Trujillo
California State Polytechnic University Pomona

Abstract

Respiratory diseases constitute a significant portion of global mortality, with COVID-19 highlighting their rapid and devastating impact. This report explores the challenges associated with diagnosing respiratory diseases, particularly in light of the uncertainties and the growing reliance on online medical resources for information. Through text mining techniques and machine learning models, the project aims to automate the interpretation of complex medical conversations and optimize the diagnosis process. The research underscores the importance of effective patient-doctor communication and the role of technology in improving diagnostic accuracy and patient outcomes. By integrating text mining into medical practice, healthcare professionals can expedite decision-making, enhance diagnostic accuracy, and optimize treatment plans. However, challenges such as data security and algorithm validation must be addressed to realize the full potential of text mining in healthcare.

Keywords: Text Analysis, TF-IDF, Predictive Labelling, Classification Analysis, Respiratory Diseases

1 Introduction

Accounting for about 20% of the global proportional mortality, respiratory diseases affect a large part of our daily lives whether it be acute or chronic type. COVID-19 played its role in reminding us how quick and destructive a respiratory disease can be on a worldwide scale. This was due to the rapid increase in varying risk factors such as population growth and urbanization – leading to closer and more frequent interpersonal contact. On the other hand, other respiratory diseases or disorders tend to have an associated risk factor related to economic growth and regional industrialization, where there is an increase in Air Quality Index (AQI) through atmospheric pollution. As the list of respiratory related diagnoses becomes extensive, it is common to see similar pre-indications such as difficulty breathing, fever, chest pain, and cough. Upper and lower ARIs (Acute Respiratory Infections), including pneumonia, are frequent at all ages but are more devastating in young children. COPD, asthma, Tuberculosis, and lung cancer are all the leading causes of respiratory morbidity and mortality among adults.

With the recency of COVID-19, many other respiratory diseases can come with its uncertainties and can invoke a sense of concern across many individuals. Being in a state of anxiousness, many try to answer their medical queries through online forums, chatbots, emails or some other form of communication as a proactive measure; however, it should be noted that this should not take the place of visiting a doctor when the symptoms become severe. Through the different channels in which medical conversations between healthcare providers and patients can occur, many patients might express dissatisfaction through the lens of poor provider communication. More meaningful communication is usually associated with improved health outcomes, as well as better patient rapport, reduced health service utilization, and ultimately better

patient and provider outcome. It is important to be aware of how these issues can be addressed in order to optimize the patient-doctor communication process. Depending on the channel of information, the process can be crucial towards working to fix these issues.

As this project leverages transcribed, simulated conversations, the use of text mining will allow us to take the volume of the data and automate the interpretation of a complex dialect. The process allows for a pipeline of translating text into key features quantitatively. Focusing on identifying which respiratory disease is on the table, we must look further into studying the symptoms and diagnosis of these diseases. From there, we would be able to manage and improve our understanding of these conditions and mitigate uncertainty for the resulting diagnosis. Within our project, we take a look at simulated conversations between patients and doctors. Within these conversations, we can extract any key information such as medical terminology indicating any important implications for the observed diseases. We then show how machine learning can optimize the latency in providing a diagnosis. Our model was trained using a quantitative representation of the patient-doctor corpus, and was able to interpret similarities across conversations in order to provide a multi-label classification output.

The value of this project comes from our goal, which is to analyze and interpret medical conversations, then providing a list of 3 diseases connecting to what is expressed. Identifying the most common diseases discussed in these conversations, we can gain insight towards the most prevalent health issues, including symptoms and appropriate treatments for these diseases. Further study of this topic can potentially lead to optimizing patient flow in a medical setting, leading to quality patient experiences. While this is not to replace the role of a healthcare professional, it serves as a tool to alleviate the uncertainty presented in some situations such as eliminating

disease possibilities. This project aims to highlight the crossover between healthcare and technology, and provide more insight towards the future of how combining these two fields can lead to improved patient satisfaction and clinic-flow efficiency. This research can also open up more opportunities for transcribing auditory data from patient-doctor conversations as it can feature scalability for larger populations and demographics.

2 Literature Review

"Modern Clinical Text Mining: A Guide and Review" (Percha, 2021) describes how Electronic health records (EHRs) are becoming a vital source of data for healthcare quality improvement, but most of the information is buried in unstructured text. It's related to the topic because it discusses how modern clinical text mining systems have accomplished a great deal, including being able to tag various clinically-relevant entities in text, how it is able to map them to standard concepts from ontologies and lexicons, and how it is able to detect uncertainty and negation, as well as understanding the person or people to whom they prefer. (Percha, 2021, p. 17)

"Text mining in long-term care: Exploring the usefulness of artificial intelligence in a nursing home setting" (Hacking et. al., 2022) talks about how narrative data within nursing homes is collected to evaluate the quality of the care that residents or their family members are perceiving. This is initially seen as a large amount of textual data. However, when the amount of data increases, the human capability to analyze it is reduced, thus using text mining approaches to evaluate the data, which is related to the topic. The result of this study showed that throughout the interviews, residents, family members, and health care professionals spoke 285, 362, and 549 per interview. (Hacking et. al., 2022, p.5) When using a word frequency

analysis, words that occurred most commonly were often positive. It also showed that care professionals had expressed a more diverse sentiment in comparison to family members and residents. Overall most interviews displayed a neutral sentiment.

"Text mining in healthcare. Applications and opportunities" (Raja et. al., 2008) discusses how the capabilities of text mining can be harnessed within healthcare settings, which is similar to our topic. Their research initially led them to believe that text mining can be an effective tool within healthcare datasets due to the shift in electronic records, as well as the availability of standardized vocabulary. Due to the extensive amount of use with vocabularies collected by the National Library of Medicine in the UMLS system, text mining is believed to be an effective choice in healthcare analytical projects (Raja et. al., 2008, p.6). "A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining" (Islam et. al., 2018) had a similar idea, highlighting that there is a lot of useful data in the growing healthcare industry, including patient demographics, treatment plans, and insurance coverage. This can all be used by both clinicians and scientists alike. However, the authors mention how there is a lack of a systematic and comprehensive narrative on the topic. They came up with a conclusion that the availability of big data created a promising research avenue for practitioners and academicians. As mentioned in their article, there is an increased number of publications in recent years which corroborates the importance of healthcare analytics, which improves healthcare systems worldwide. This is related to the topic because they also mention how the goal is to facilitate coordinated and well-informed healthcare systems capable of ensuring maximum patient satisfaction (Islam et. al., 2018, p.35).

"Portable automatic text classification for adverse drug reaction detection via multi-corpus training" (Sarker et. al., 2015) relates to the topic because they mention how they are collecting features from text. The topic of their article has to do with the

Automatic detection of adverse drug reaction (ADR). However they involve sentences from clinical reports, but mainly social media. By utilizing NLP techniques, they found out that these techniques can be applied to improve the automatic classification of social media text containing medical information (Sarker et. al., 2015, p.205). With "A text mining approach to the prediction of disease status from clinical discharge summaries" (Yang et. al., 2009) the focus is to identify the status of obesity and 15 related co-morbidities in patients using their clinical discharge summaries. This is related to our topic because they had a textual task, which was to identify explicit references to the diseases. They also had an intuitive task, which was to predict the disease status when the evidence was not explicitly asserted. The findings resulted in the performance being in line with the agreement between human annotators. This means that there is potential for text mining when it comes to accurate and efficient prediction of disease statuses from clinical discharge summaries (Yang et. al., 2009, p.596).

"Text Mining for Precision Medicine: Bringing structure to EHRs and biomedical literature to understand genes and health" (Simmons et. al., 2016) involves answering the question of whether it is possible to find clinically actionable granularity in diagnosing disease and classifying patient risk. The article relates to our topic due to the similarity in diagnosing diseases. The authors came up with the conclusion that text mining is a vehicle used to obtain increased utility from existing information resources. It also offers several advantages in the precision medicine value equation. An example would be mining biomedical literature, which allows for streamlined curation, and improved research efficiency through hypothesis generation (Simmons et. al., 2016, p.18). The next article listed is called "Overcoming barriers to NLP for clinical text: the role of shared tasks and the need for additional creative solutions" (Chapman et. al., 2011). This particular article highlights the barriers of NLP de-

velopment in the clinical domain. They also make note that these barriers also pop up in software engineering and in general NLP. This is related to our topic due to the relation of patient information. However there are concerns of patient privacy and revealing unfavorable institutional practices, which is causing clinics and hospitals to become reluctant to allow access to clinical data. (Chapman et. al., 2011 p. 540). The evaluations they made in their research contributed to vitalizing the field of clinical NLP but mention how critical issues of data access need to be addressed. "Extracting information from textual documents in the electronic health record: a review of recent research" (Meystre et. al., 2008) talks about the rapid adoption of Electronic Health Records (EHR), as well as the need for improved quality and reduced medical errors being strong incentives for the development of Natural Language Processing (NLP). This is related to our topic due to the similarity of using NLP with our data. The conclusion states that performance has gradually improved, and that systems are mostly statistically based. However, creating annotated clinical text corpora is one of the main challenges for the future of this field (Meystre et. al., 2008, p.141).

Finally, our last article "SemEHR: A general-purpose semantic search system to surface semantic data from clinical notes for tailored care, trial recruitment, and clinical research" (Wu et. al., 2018) has to do with trying to unlock data within both structured and unstructured components of electronic health records (EHRs). This is related to our topic of going through data related to clinical records. The authors found that using these methods will provide a step change in data available to use for multiple purposes like hospital management and trial recruitment. They used an implemented SemEHR, which is an open source semantics search and analytics tool for EHRs and found that SemEHR turns IE tasks into ontology-based searches. This

significantly lowers the barriers to secondary use of unstructured EHR data, and is deployed in several NHS hospitals in the UK (Wu et. al., 2018, p. 537).

3 Data

For the purpose of this project, our sources of data and information were distinct entities such that one was the compilation of simulated conversations between patients and doctors while another was a constructed table detailing key information about the respective disease/disorder.

3.1 Data Sources

For the compilation of simulated conversations, the data was obtained from an open-access medical journal (12) primarily focusing on respiratory diseases. Originally, there were 272 simulated conversations across 5 medical domains; however, making up the bulk of the folder, 213 files were related to a respiratory disease. Each text file in the folder represents a unique patient-doctor dialogue in the format of Objective Structured Clinical Examinations (OSCE), manually checked for any quality or integrity concerns to its corresponding audio recording. As mentioned in the article as well, "the 'physician' was blinded to the final diagnosis to simulate the clinic and hospital setting, and to avoid asking leading questions". Thus, it is apparent, measures were taken to avoid any bias in the development of the conversation – introducing room for potential error and uncertainty. For the constructed table, we collected key information from Mayo Clinic, a credible source for disease-related information. For 19 unique respiratory diseases/disorders, the following information was collected on the respective webpage: Disease Name, List of Symptoms, List of Treatments, and

a Summary. This collection of information provides more context to each disease, marking similarities or highlighting differences between them.

3.2 Data Pre-Processing

Prior to feeding the data into the methodologies that follow, a few pre-processing techniques were required of the text for proper analysis. For each text file, the lines were indicative of the person (i.e. patient (P) or doctor(D)) talking in the dialogue, tending to be alternating responses. As it was not needed, the identification of the person speaking was removed and each line was then concatenated. For both documents, it was needed that the corpus was filtered of stop words. In our situation, we decided on 3 concatenations of stop-words lists. The first was the standard 'nltk' package in which common English words were removed. The second was a clinical stop words list (13), removing non-unique medical terminology. And the last list was a manually created one in which words that were not caught by either due to randomness introduced by humans. In these processes, the words were lowercased and tokenized for easier comparisons.

4 Methodology

Following the pre-processing procedures on the transcripted conversations, the applied methodologies in this section are exec on the basis of a numerical representation of the text. Thus, the approaches used for a computer to interpret the conversations were done in the following order: Term Frequency - Inverse Document Frequency (TF-IDF), Predictive Labeling, and Multi-Label KNN.

4.1 Term Frequency - Inverse Document Frequency

Term Frequency - Inverse Document Frequency (or TF-IDF) is a statistical method that measures how important a term is relative to a corpus. Thus, a higher frequency value will indicate the importance of a term while the opposite occurs for a lower frequency value. In the proposed version, we ignored terms that appeared in less than 30% and more than 80% of the documents. In doing so, this isolates words that are more common across the conversations but still relative to the topic of respiratory diseases/medical field. Setting these thresholds to 30% and 80%, we can avoid jargon and repetitive words non-specific to a particular disease, respectively. In addition, the proposed version also explores the variations of unigrams and bi-grams, providing opportunities for joint words to have meaning such as "high fever" or "low energy".

4.2 Predictive Labeling

Given that the dataset does not explicitly provide a respiratory disease/disorder in the form of a label, it was imperative that a computational procedure was developed to provide each conversation with a Top 3 list of expected diseases/disorders based on a calculated score. For each conversation, every disease in our constructed list was considered as an equal candidate. Then, for each word in our TF-IDF vocab list score1,score2, score3, and score4 were calculated as the sum of the weights for each time the word appeared in the tokenized list of Disease, Symptoms, Treatment, and Summary, respectively. For each disease, a final score was derived as the sum of its individual sums for which the disease labels would be selected corresponding to the top 3 scores.

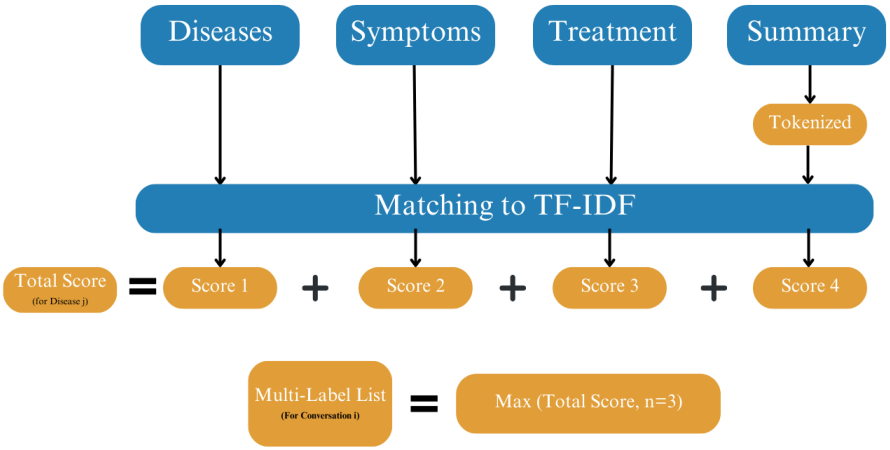

Figure 1: Flow Chart for Labeling Conversations

4.3 Multi-Label K-Nearest Neighbors

Aiming to give multiple possibilities of a diagnosis due to human uncertainty, a machine learning model can be used for classification purposes. Given that the data has been pre-processed to be interpreted through numerical values, KNN can perform through the computations of Euclidean distances between its k=10 closest neighbors and predict based on the majority. In this particular variation, we are dealing with a multi-label KNN, meaning that we provide a one-hot encoded representation of the top 3 diseases for each conversation through a list of our 19 explored diseases and the model will output another one-hot encoded representation of which diseases it predicted.

5 Analysis and Results

From the testing set, it is highly probable that accuracy might no longer be a proper measure in this classification setting from the imbalance of the few observations tested on. Unlike the binary or multi-class classification problems, the classes in multi-label

classification problems are not mutually exclusive. Thus, rather than penalizing the model for not providing an exact match, we relaxed the metrics to account for any matching disease between true and predicted labels for each conversation.

The model results in Figure 2 provide a classification report on the test set. For some context, precision makes up for the accuracy of the positive predictions in the model. Recall (or sensitivity) measures the ability of the model to identify all the relevant samples. The F1-score in the dataset represents precision/recall and provides a single metric that combines both precision and recall into one value making it a useful metric when you want to balance between precision and recall.(F1-score is calculated as $2\times\frac{precision\times recall}{precisionrecall}$). Support in the dataset refers to the number of actual occurrences of the class in your dataset (i.e. the number of instances of asthma). Breaking down the averages, micro average, represents metric calculations counting the total true positives, false negatives, and false positives in this case micro weight for each instance. Macro average calculates metrics independently for each class and then takes the unweighted mean this gives equal weight for each instance. Weighted average measures the average weighted by support, in the report the average considers the class imbalance by metrics support. Sample averages consider each of the individuals' predicted measures equally. Thus, looking into Figure 2, we can observe Pnuemothorax and Cystic Fibrosis having the value of 0, indicating misrepresentation in the test set or poor performance. However, for other diseases such as strep throat, we are getting a precision score of 1.00, a recall score of 0.71, and a F1-score of 0.83, indicating an overall good performance for the 7 cases.

Another metric used to evaluate the classification performance of the model was Hamming Loss. The Hamming Loss metric lessens the penalization between the true and predicted labels such that it allows for partially correct predictions. For example, say the model was able to predict one of the 3 true labels then the fraction of remaining

	precision	recall	f1-score	support
acute respiratory distress syndrome	0.67	0.29	0.40	7
asbestosis	1.00	0.28	0.43	18
aspergillosis	0.00	0.00	0.00	0
asthma	0.78	0.58	0.67	12
bronchiolitis	0.00	0.00	0.00	0
bronchitis	0.00	0.00	0.00	0
chronic bronchitis	0.00	0.00	0.00	0
chronic obstructive pulmonary disease	0.75	0.18	0.29	17
influenza	0.50	0.11	0.18	9
pneumonia	1.00	0.20	0.33	5
pneumothorax	0.00	0.00	0.00	1
respiratory syncytial virus	0.00	0.00	0.00	1
tuberculosis	0.00	0.00	0.00	3
bronchopulmonary dysplasia (bpd)	0.00	0.00	0.00	1
cystic fibrosis	0.00	0.00	0.00	0
pulmonary fibrosis	0.53	1.00	0.70	23
sleep apnea	0.00	0.00	0.00	3
covid	0.61	1.00	0.76	22
strep throat	1.00	0.71	0.83	7
micro avg	0.64	0.53	0.58	129
macro avg	0.36	0.23	0.24	129
weighted avg	0.67	0.53	0.51	129
samples avg	0.65	0.53	0.58	129

Figure 2: Model Metrics

incorrect labels over the total number of labels would count towards the Hamming Loss. In Figure 3, we can see the Hamming Loss for each conversation in the testing set where 0.231052631578947367 is the Hamming Loss for the first conversation and 0.15789473684210525 is the Hamming Loss for the second conversation and so on. These values suggest the proportion of incorrectly predicted labels for that given conversation in the test set, so 23.11% and 15.79%, respectively. On average for the 43 observations in the test set, we get an average Hamming Loss of 0.12117503059975519, or on average 12.12% of the instances were incorrectly predicted by the classifier. We can observe that one instance has a Hamming Loss of 0.0, which is the ideal value as it indicates each label was predicted correctly for the particular observation. The overall measure of the performance of how it handles the multi-label classification for respiratory cases, so lower values indicate better performance. While there may only be 43 observations in the test set there are 129 total labels to account for as each conversation has 3 associated diseases, thus 12.12% suggests good predictive performance from the model.

```
Individual Hamming Loss:
 [0.21052631578947367, 0.15789473684210525, 0.05263157894736842, 0.15789473684210525, 0.10526315789473684, 0.157894
73684210525, 0.21052631578947367, 0.15789473684210525, 0.15789473684210525, 0.15789473684210525, 0.2105263157894736
7, 0.05263157894736842, 0.21052631578947367, 0.10526315789473684, 0.05263157894736842, 0.10526315789473684, 0.10526
315789473684, 0.10526315789473684, 0.15789473684210525, 0.0, 0.2631578947368421, 0.10526315789473684, 0.15789473684
210525, 0.15789473684210525, 0.05263157894736842, 0.05263157894736842, 0.15789473684210525, 0.10526315789473684, 0.
10526315789473684, 0.05263157894736842, 0.10526315789473684, 0.10526315789473684, 0.10526315789473684, 0.1052631578
9473684, 0.05263157894736842, 0.10526315789473684, 0.05263157894736842, 0.05263157894736842, 0.15789473684210525,
0.2631578947368421, 0.15789473684210525, 0.05263157894736842, 0.05263157894736842]

Average Hamming Loss:
 0.12117503059975519
```

Figure 3: Hamming Loss of KNN model

6 Discussion

Implementing text mining to find descriptors linked with respiratory diseases/disorders using TF-IDF and classification models could have a substantial influence on the healthcare system. One of the most significant advantages is the potential reduction in the time it takes for patient information to reach the doctor, especially in circumstances where urgency is critical. By analyzing transcribed discussions between patients and doctors, the research speeds up the process of finding relevant symptoms and their links to respiratory disorders. This not only allows patients to get insight into prospective disorders, but it also gives clinicians a more complete understanding of the patient's health, allowing them to make more educated and timely judgments. Furthermore, by providing a top N list of potential diagnoses based on the identified descriptors, the initiative improves diagnostic efficiency, shortening the time from symptom presentation to treatment initiation. In addition to reducing the time it takes for patient information to reach the doctor and speeding up the diagnosis process, text mining deployment can have a number of other benefits for the healthcare ecosystem. To begin, by using transcribed conversations to discover descriptors linked with respiratory disorders, the project enables the establishment of a comprehensive knowledge base that can be continuously updated and developed over time. This knowledge base can be a significant resource for medical practitioners, giving them access to a plethora of information about symptoms, illness patterns and treatment

outcomes. Furthermore, by utilizing machine learning methods such as classification models, the project allows for the automatic detection and categorization of symptoms, lowering the strain on healthcare personnel while enhancing diagnostic accuracy and consistency. Additionally, as the model is improved it can be used for other fields of medicine such as gastrointestinal diseases.Overall, this initiative has the potential to improve patient outcomes and healthcare delivery by utilizing advanced text mining algorithms to speed up the diagnostic process for patients and doctors.

7 Conclusion

Finally, the application of text mining in medical care is a significant step forward, with far-reaching implications for patient outcomes, healthcare delivery, and research. Text mining enables healthcare professionals to accelerate decision-making, enhance diagnosis accuracy, and optimize treatment plans. It makes it easier to develop large knowledge bases that can be used for evidence-based practice, clinical decision support systems, and medical research. However, it is critical to recognize the challenges associated with implementing text mining in medical care, such as the need for strong data security safeguards to protect patients' privacy. It is also important to constantly validate and refine algorithms to insure accuracy with diagnoses. Even then, text mining is a tool that doctors can utilize to support their thoughts or point out diseases they had not considered. Despite these obstacles, text mining integration has the potential to change healthcare delivery, empower patients, and ultimately improve health outcomes for people all over the world. As the area evolves, interdisciplinary collaboration among data scientists, healthcare practitioners, and policymakers will be critical to realizing text mining's full potential to revolutionize the future of medical care.

References

[1] Practical Approach to Lung Health: Manual on Initiating PAL Implementation. Geneva: World Health Organization; 2008. 2, Estimating the burden of respiratory diseases. Available from: https://www.ncbi.nlm.nih.gov/books/NBK310631/

[2] Percha B. Modern Clinical Text Mining: A Guide and Review. Annu Rev Biomed Data Sci. 2021 Jul 20;4:165-187. doi: 10.1146/annurev-biodatasci-030421-030931. Epub 2021 May 26. PMID: 34465177.

[3] Hacking, C., Verbeek, H., Hamers, J. P. H., Sion, K., & Aarts, S. (2022). Text mining in long-term care: Exploring the usefulness of artificial intelligence in a nursing home setting. PloS one, 17(8), e0268281. https://doi.org/10.1371/journal.pone.0268281

[4] Raja, Uzma & Mitchell, Tara & Day, Timothy & Hardin, James. (2008). Text mining in healthcare. Applications and opportunities. Journal of healthcare information management : JHIM. 22. 52-6.

[5] Islam, M. S., Hasan, M. M., Wang, X., Germack, H. D., & Noor-E-Alam, M. (2018). A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining. Healthcare (Basel, Switzerland), 6(2), 54. https://doi.org/10.3390/healthcare6020054

[6] Sarker, A., & Gonzalez, G. (2015). Portable automatic text classification for adverse drug reaction detection via multi-corpus training. Journal of biomedical informatics, 53, 196–207. https://doi.org/10.1016/j.jbi.2014.11.002

[7] Yang, H., Spasic, I., Keane, J. A., & Nenadic, G. (2009). A text mining approach to the prediction of disease status from clinical discharge summaries. Jour-

nal of the American Medical Informatics Association : JAMIA, 16(4), 596–600. https://doi.org/10.1197/jamia.M3096

[8] Simmons, M., Singhal, A., & Lu, Z. (2016). Text Mining for Precision Medicine: Bringing Structure to EHRs and Biomedical Literature to Understand Genes and Health. Advances in experimental medicine and biology, 939, 139–166. https://doi.org/10.1007/978-981-10-1503-8_7

[9] Chapman, W. W., Nadkarni, P. M., Hirschman, L., D'Avolio, L. W., Savova, G. K., & Uzuner, O. (2011). Overcoming barriers to NLP for clinical text: the role of shared tasks and the need for additional creative solutions. Journal of the American Medical Informatics Association : JAMIA, 18(5), 540–543. https://doi.org/10.1136/amiajnl-2011-000465

[10] Meystre, S. M., Savova, G. K., Kipper-Schuler, K. C., & Hurdle, J. F. (2008). Extracting information from textual documents in the electronic health record: a review of recent research. Yearbook of medical informatics, 128–144.

[11] Wu, H., Toti, G., Morley, K. I., Ibrahim, Z. M., Folarin, A., Jackson, R., Kartoglu, I., Agrawal, A., Stringer, C., Gale, D., Gorrell, G., Roberts, A., Broadbent, M., Stewart, R., & Dobson, R. J. B. (2018). SemEHR: A general-purpose semantic search system to surface semantic data from clinical notes for tailored care, trial recruitment, and clinical research. Journal of the American Medical Informatics Association : JAMIA, 25(5), 530–537. https://doi.org/10.1093/jamia/ocx160

[12] Smith, Christopher William; Fareez, Faiha; Parikh, Tishya; Wavell, Christopher; Shahab, Saba; Chevalier, Meghan; et al. (2022). A dataset of simulated patient-physician medical interviews with a focus on respiratory cases. figshare. Collection. https://doi.org/10.6084/m9.figshare.c.5545842.v1

[13] Ganesan K, Lloyd S, Sarkar V. Discovering Related Clinical Concepts Using Large Amounts of Clinical Notes. Biomed Eng Comput Biol. 2016 Sep 7;7(Suppl 2):27-33. doi: 10.4137/BECB.S36155. PMID: 27656096; PMCID: PMC5015701.