# Introduction

Radiology reports are used by departments internationally in order to provide patients with vital information regarding their diagnosis. Nevertheless, a common theme among reports is a lack of intelligible dialogue and an unfocused style leading to clinically futile information (1). While the impressions section of report has become a common theme in radiology that can provide summary-like information to the patient which includes their symptoms, findings, official medical diagnosis, and future needed steps for the radiology department to complete (2), there is no definitive measure of distance between multiple reports and their findings. Because many reports are formatted in similar styles (3), any measure of similarity between 2 reports must filter through vital and repeated information. This question has been explored in radiology (4) through a lens of the ct-scans and image classification, however, one that uses the language found in the reports to measure distance has not been attempted (5).

Natural language Processing (NLP) has been previously applied to create vector distances between bodies of text (6). This task is done through the use of word embeddings – universalized formatting for words to be represented and compared with one another (7). The practice of radiology embeddings has been applied with common databases for radiology terms such as radiopaedia (8); however, embeddings have not been comparatively used with reports to produce a similarity index.

In our work, we show that embeddings between words can be used to create a distance score between radiology reports. We use the labels (9) and article titles (10) on radiopaedia – a database of radiology terms and their medical significance (11) – to find matches between 2 reports. [ADD NLP findings here]

References (not properly cited yet – ask Dr. Sohn about type of formatting) - Zotero:

1. <https://www.jacr.org/article/S1546-1440(18)30712-9/fulltext#relatedArticles>
2. <https://www.radiologyinfo.org/en/info/article-read-radiology-report>
3. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6472864/>
4. <https://www.researchgate.net/post/What_are_the_best_methods_to_compare_the_similarity_of_two_medical_images>
5. <https://www.sciencedirect.com/science/article/pii/S1532046413000889>
6. <https://www.baeldung.com/cs/ml-similarities-in-text>
7. <https://machinelearningmastery.com/what-are-word-embeddings/>
8. <https://www.sciencedirect.com/science/article/abs/pii/S1532046420302938?via%3Dihub>
9. <https://radiopaedia.org/encyclopaedia/classifications/all?lang=us#collapse-by-system>
10. <https://radiopaedia.org/articles/article-title?lang=us>
11. <https://radiopaedia.org/about?lang=us>

Notes:

* 3 paragraphs with 1 (current setup + background info). 2 (other work in the field that is related and how it can be helpful), 3 (hypothesis and contribution in your project)
* <https://link.springer.com/article/10.1007%2Fs10278-020-00348-8>