SCORING PATIENT NOTE TAKING USING NLP

Ritika Singh | Prajjwal Gupta | Saurabh Singh | Dr. R.Sathyaraj | School of Computer Science & Engineering

Introduction

We propose a methodology for National Board of Medical Examiner, which accesses the skills of writing patient's notes for Medical Licensing Examination. The process of assessing the notes for every candidate manually is very time consuming for the trained physicians. Using NLP, the task of identifying clinical concepts in patient's notes following the exam rubric will be done.

Motivation

Almost 90% of the 2.5 quintillion bytes of data that is being produced each day is unlabelled and unuseful. Patient's notes and clinical records of these candidates manually by the trained physicians requires significant time along with human and financial resources

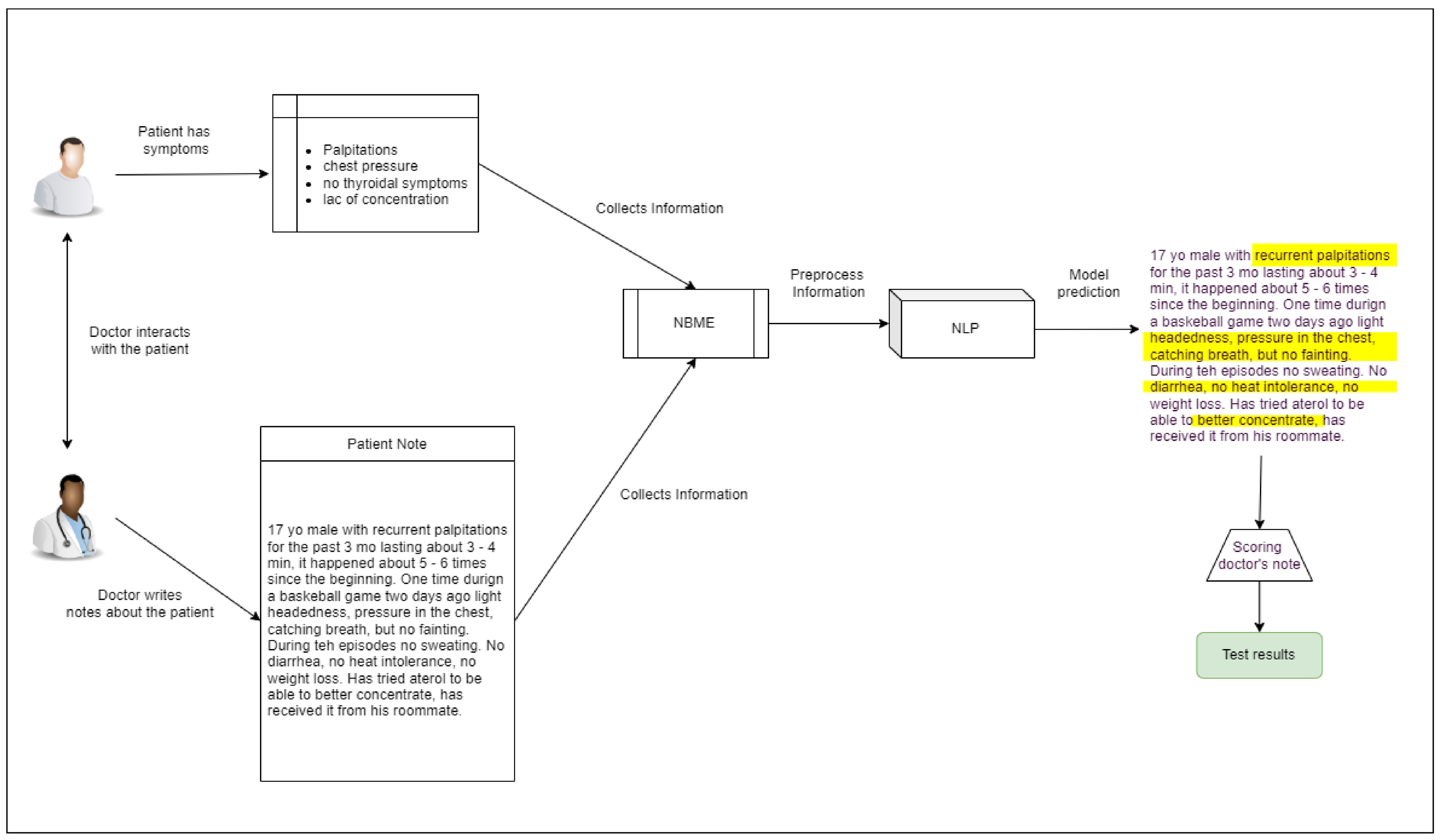
SCOPE of the Project

1- Automation : Automate the manual task of trained physicians to analyze all the candidates notes to correctly map the features or diseases with the patients symptoms, problems and medical history using NLP models.

2- Handling Ambiguity with user friendly Web application which can help the assessors to score directly without wasting lot of resources.

3- Applications: Future applications include prediction medicines and diagnosis of diseases directly from notes analysis. The large amount of data can be used for better purposes on more understanding and training.

Methodology



Information Flow:

Step 1: A patient is identified who is to be assessed. The features (symptoms and conditions) evident and present in the patient is noted down.

Step 2: The doctor interacts with the patient by asking them questions related to the conditions and makes notes of the same.

Step 3: The doctor's notes along with the identified features in the patient are entered into an excel sheet (.csv file)

Step 4: The prepared file is uploaded on our web portal where it is processed and the trained models carry out their intended tasks.

Step 5: The final results returned by the model are displayed on the webpage.

The given problem statement can be posed as a combination of test segmentation and question answering task.

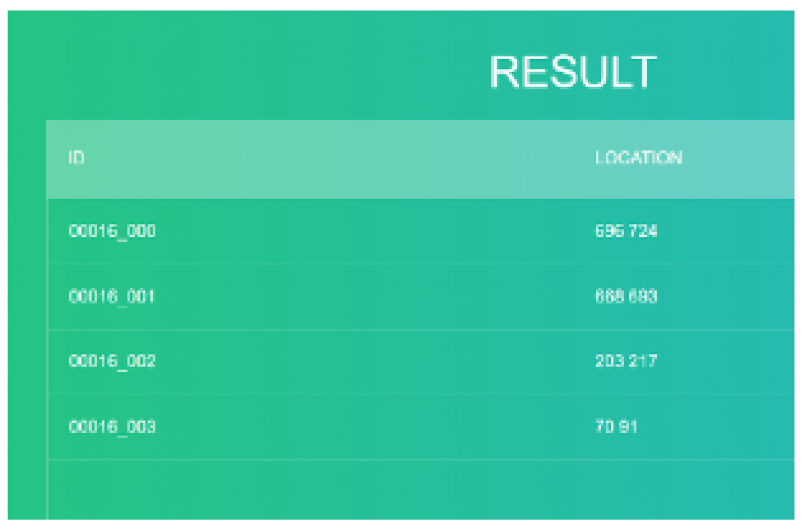
Text segmentation: The process of splitting written text into meaningful components, such as words, sentences, or subjects, is known as text segmentation. Natural language processing refers to both the mental processes that people utilise while reading text and the artificial processes that are implemented in computers. In our project the main output are the text segments in the doctor's notes which have identified particular features present in the patient.

Question Answering: Question answering is a major NLP challenge as well as a long-standing AI milestone. A user may ask a question in plain language and receive an immediate and concise response using QA technologies. The capacity to read a piece of literature and then answer questions about it is known as reading comprehension. Reading comprehension is challenging for computers because it necessitates a combination of natural language comprehension and global knowledge. In our project, the context is provided by the doctor's notes. The questions queried are the features present in the particular patient and the answer expected is the text segment in which the doctor has identified the query feature.

Results

Results

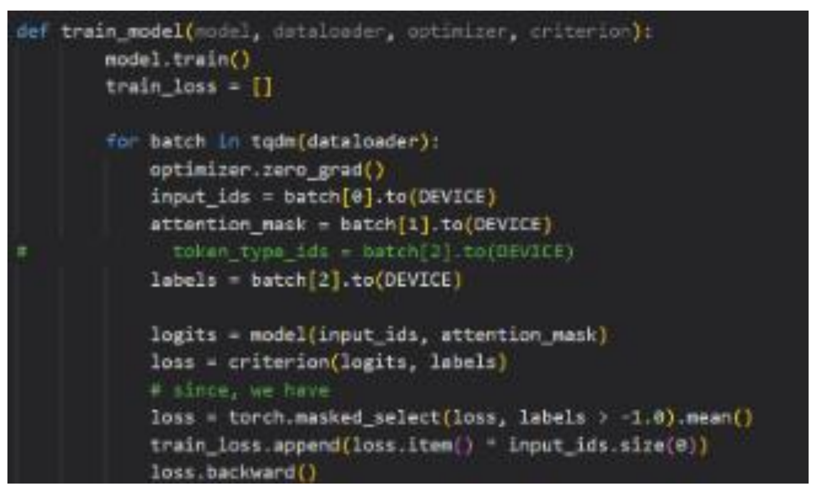
Test & Output:



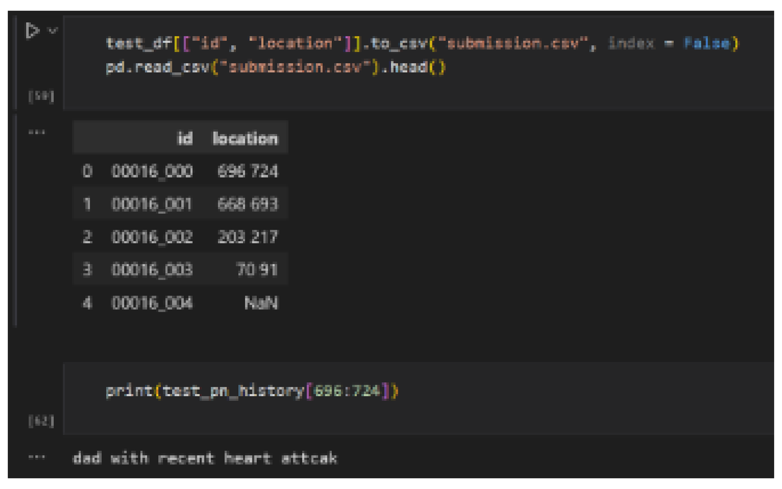
Here, we can see that the locations of the annotations describing the symptoms of the patients in the patient notes taken by the examinees are generated which can be used to calculate the results of the examinations based on how many symptoms have been correctly identified.

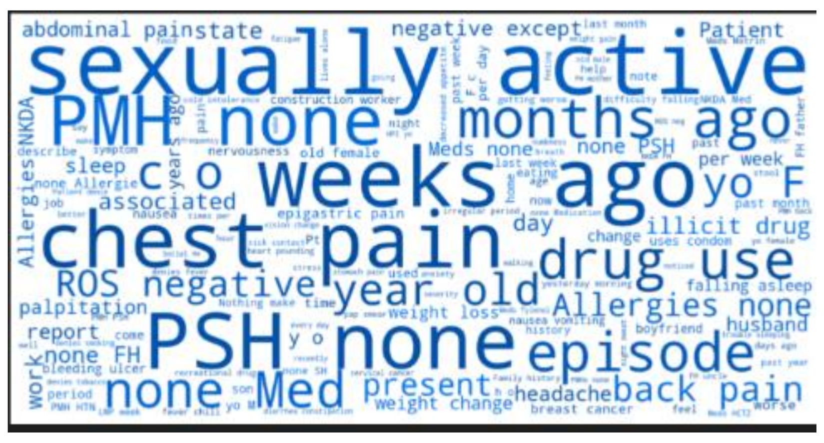
For eg: for ID 00016_000 location 696 to 724 includes "dad with recent heart attack"

Model and training:



Output: Result & Visualization





Conclusion

So, we can conclude that checking examination papers for patient note-taking can be made a lot simpler by automating the whole process. Using various Natural Language processing visualization techniques, we can visualize the annotations in an interactive way. We can even mark the NER in the patient notes to get the features and make wordclouds for them. Using various NLP techniques like tokenization we are preprocessing the dataset. Using the ROBERTA model, we are able to accurately mark the locations of the annotations i.e the symptoms of the patients who have been taken down by the doctors who are taking the exam.

So, the trained physicians who earlier had to manually check the notes, can now use the web application to upload the examination notes, and get the locations of the annotations and wether they are present or not and accordingly mark the candidates hence making the whole process a lot simpler. While testing we can see that the locations of the right annotations have been correctly marked.

References

[1] Karami, Amir & Gangopadhyay, Aryya & Zhou, Bin & Kharrazi, Hadi. (2017). Fuzzy Approach Topic Modeling for Health and Medical Corpora. International Journal of Fuzzy Systems. 20. 10.1007/s40815-017-0327-9.
[2] Rasmy, L., Xiang, Y., Xie, Z. et al. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. npj Digit. Med. 4, 86 (2021). https://doi.org/10.1038/s41746-021-00455-y
[3] Github: https://github.com/ritikaxx/Scoring-Note-Taking