



NATURAL LANGUAGE PROCESSING

Scoring patient note taking by doctors using NLP



RITIKA SINGH 19BCI0166
SAURABH SINGH 19BCI0184
PRAJJWAL GUPTA 19BCI0171

Abstract

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.

Using NLP models like BERT, ALBERTA, DEBERTA, ROBERTA we will be showing the result of the input given by the trained physicians to analyze the patient notes for all the candidates.



Importance



Almost 90% of the 2.5 quintillion bytes of data that is being produced each day is unlabelled and useless.

Patient's notes and clinical records of these candidates manually by the trained physicians requires significant time along with human and financial resources

Objective



Automation

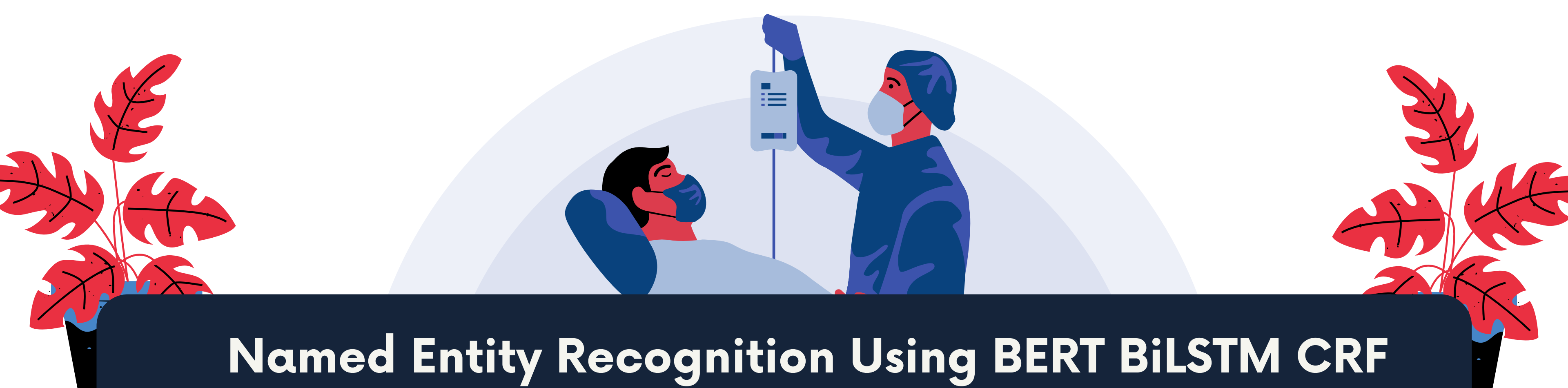
Automate the manual task of trained physicians to analyze all the candidates notes to correctly map the features or diseases with the patients symptoms, problems and medical history using NLP models.

Handling Ambiguity

Statements like "quitting job" and "no longer interested on working " referring to the same feature/ problem have to be mapped correctly according to the exam rubrics

Web Application

We will be developing a full software solution in which the input will be in the form of a csv file uploaded by the trained physicians and the output will be the mapped feature and the particular locations of the part of the notes implying the annotations for scoring the candidates.



Named Entity Recognition Using BERT BiLSTM CRF for Chinese Electronic Health Records

Z. Dai, X. Wang, P. Ni, Y. Li, G. Li and X. Bai

International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2019

This paper presents the neural network approaches to Natural Language Processing for Clinical Health Records using Named Entity Recognition. It compares various models like BiLSTM and 2 pre-trained models including word2vec and BERT. The results show that the BERT model showed the highest accuracy in extracting the valuable medical information.



A survey of word embeddings for clinical text, Journal of Biomedical Informatics

Faiza Khan Khattak, Serena Jeblee, Chloé Pou-Prom, Mohamed Abdalla, Christopher Meaney, Frank Rudzicz

**Journal of Biomedical Informatics,
Volume 100, Supplement,
2019**

This paper discusses the different types of clinical corpora, word representations, pre-trained clinical word-vector embeddings, evaluation, applications and limitations of each



An attention-based BiLSTM-CRF approach to document-level chemical named entity recognition

**Ling Luo, Zhihao Yang, Pei Yang, Yin Zhang,
Lei Wang, Hongfei Lin, Jian Wang**

**Bioinformatics, Volume 34, Issue 8, 15 April
2018**

It uses the information obtained using attention mechanism to enforce tagging consistency for multiple tokens in the same document. Features used in traditional NER methods like POS tagging are used along with these neural network models which have shown an increase in the accuracy.



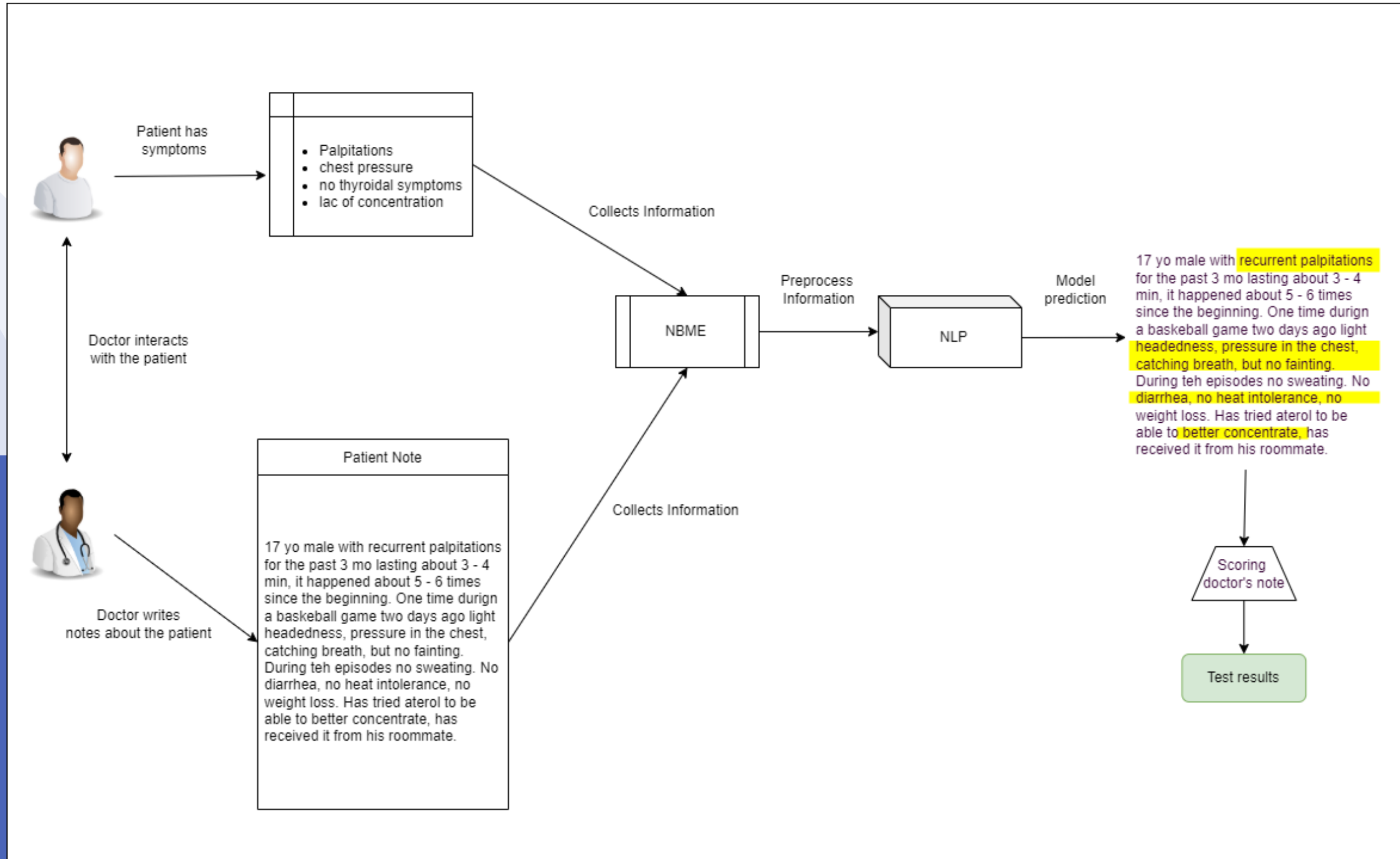
BERT-Based Transfer-Learning Approach for Nested Named-Entity Recognition Using Joint Labeling

Agrawal, Ankit & Tripathi, Sarsij & Vardhan, Manu & Sihag, Vikas & Choudhary, Gaurav & Dragoni, Nicola

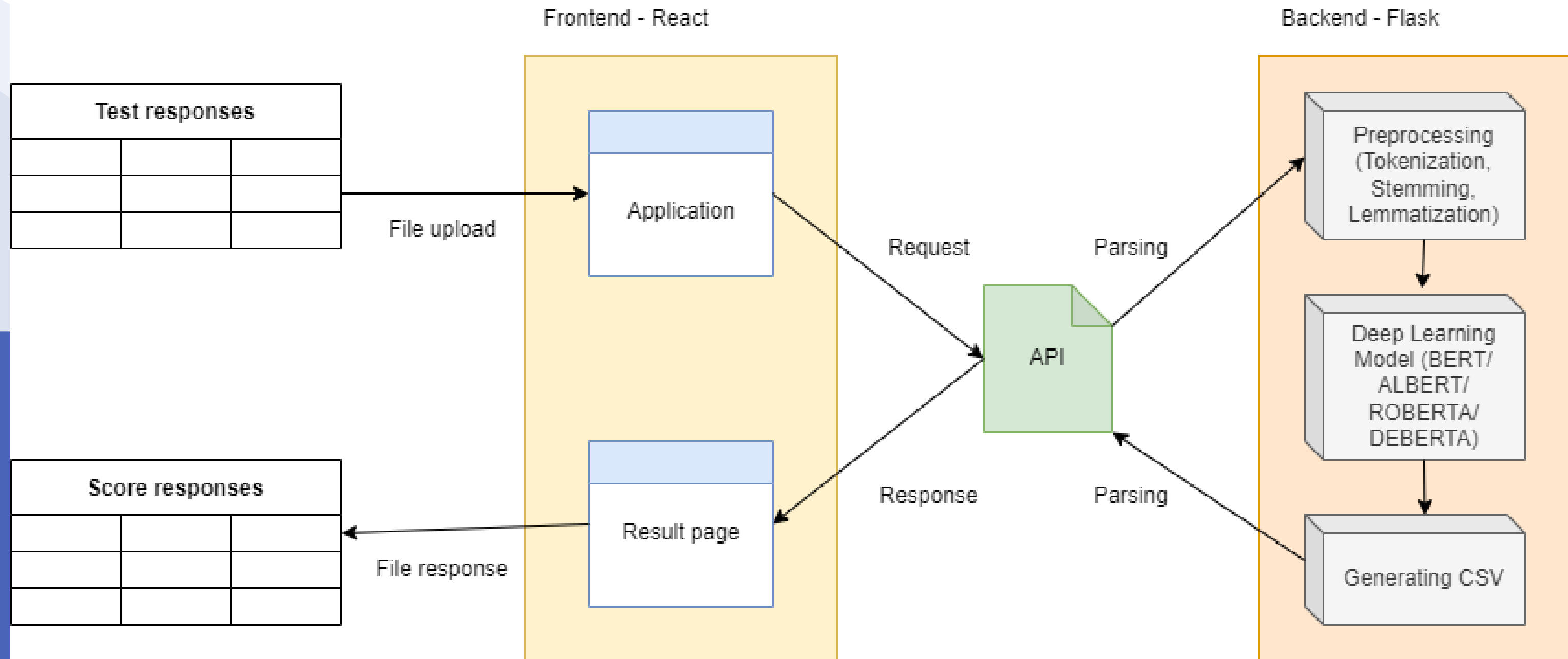
Applied Sciences. 12. 976, 2022

The authors proposed to solve the problem of nested named-entity recognition using the transfer-learning approach. Different variants of fine-tuned, pretrained, BERT-based language models were used for the problem. Two different datasets were used for four and two levels of annotations.

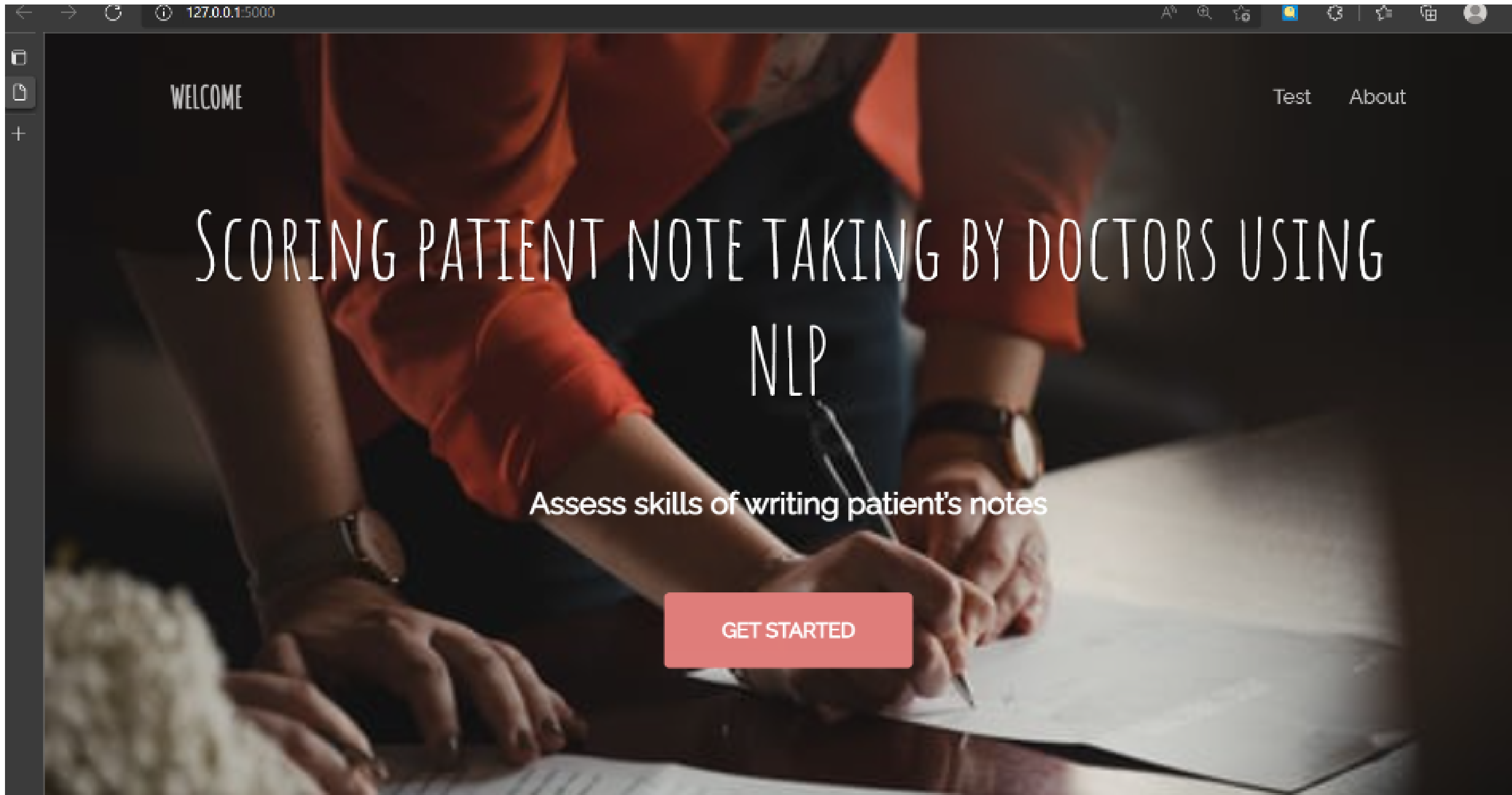
Methodology



Architecture

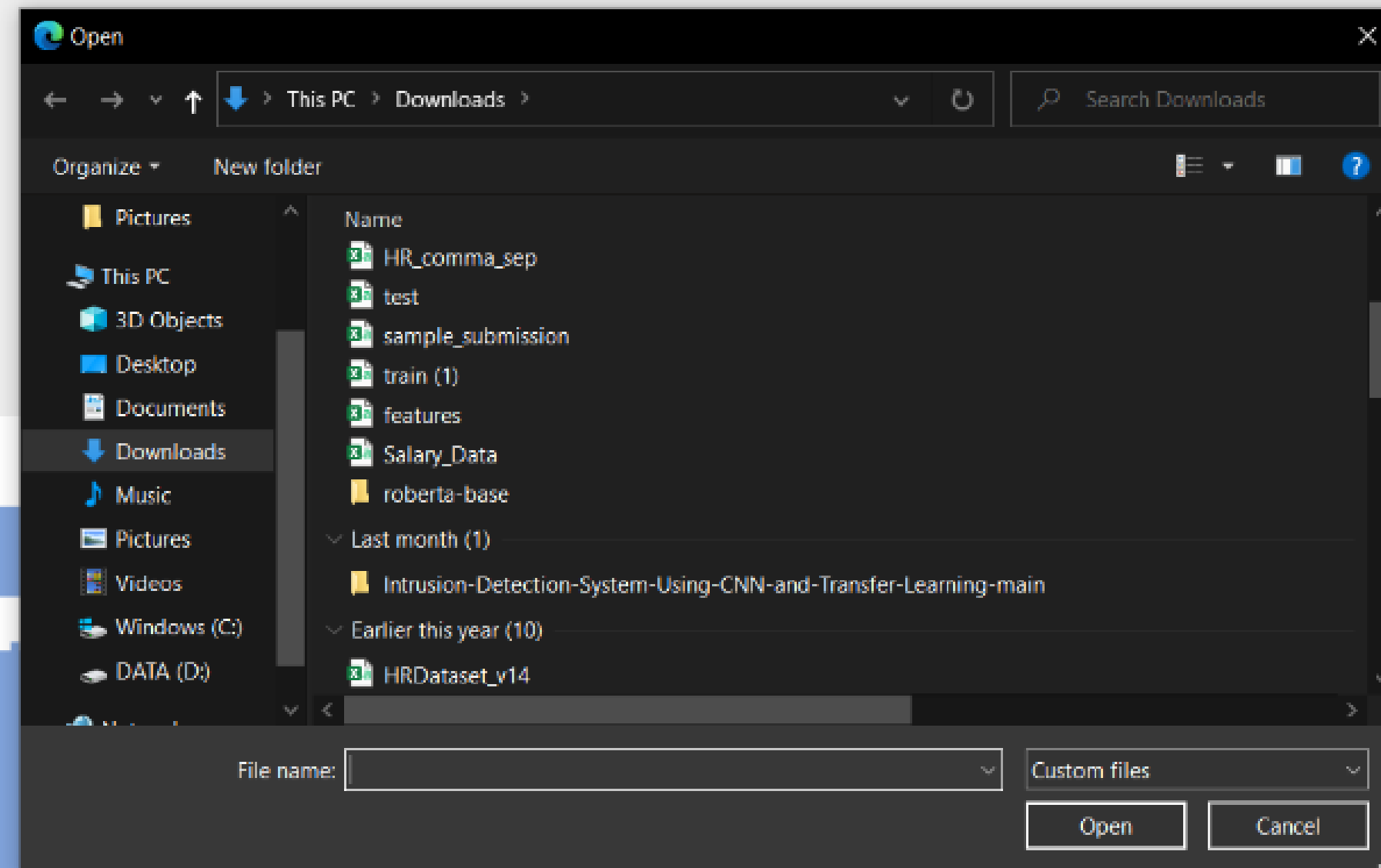


RESULTS



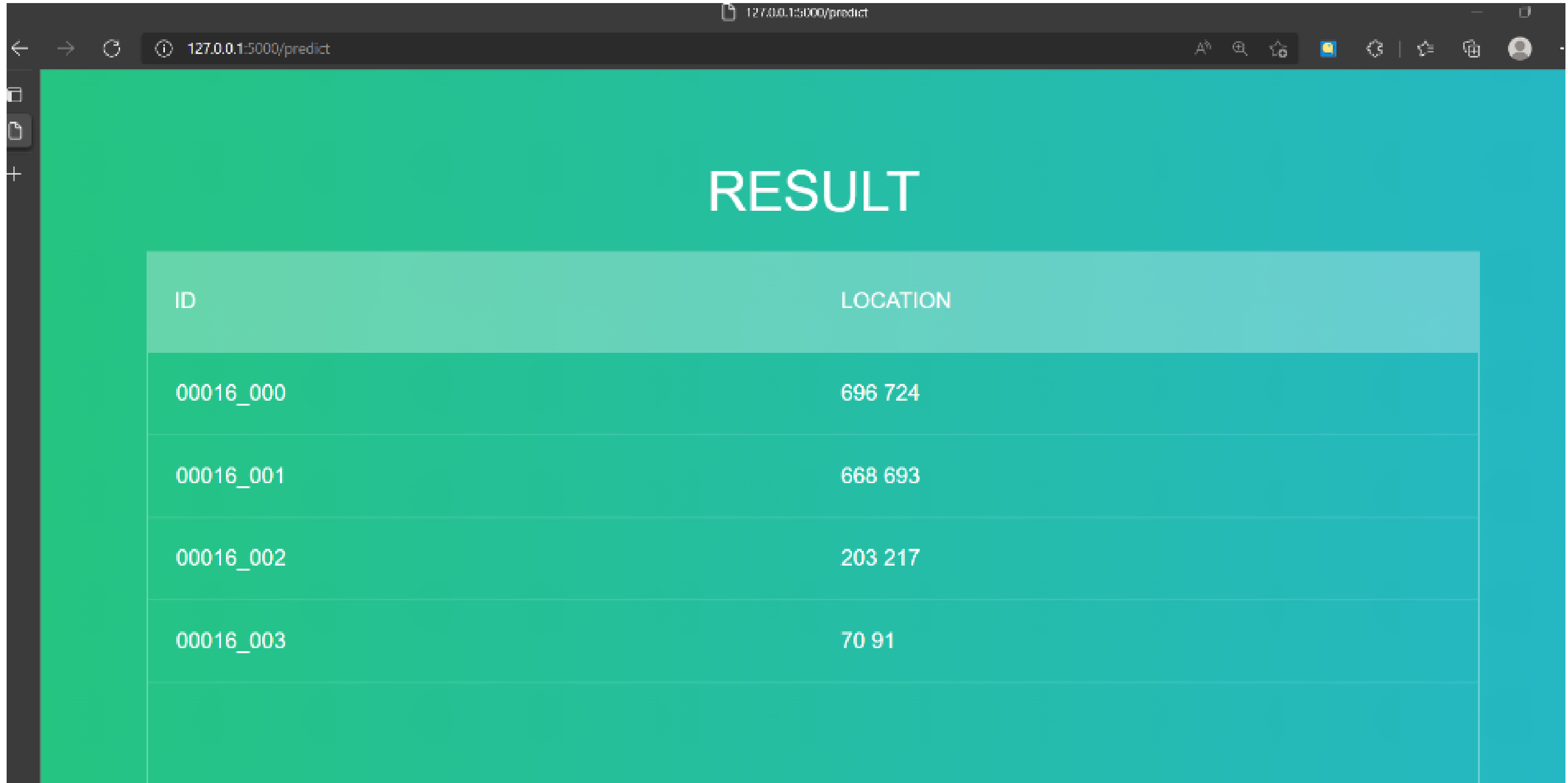
NATURAL LANGUAGE PROCESSING

RESULTS



SUBMIT

RESULTS



The screenshot shows a web browser window with the address bar displaying '127.0.0.1:5000/predict'. The page has a teal background and features a large white 'RESULT' heading. Below the heading is a table with two columns: 'ID' and 'LOCATION'. The table contains four rows of data. The browser's address bar and various icons are visible at the top of the window.

| ID | LOCATION |
|-----------|----------|
| 00016_000 | 696 724 |
| 00016_001 | 668 693 |
| 00016_002 | 203 217 |
| 00016_003 | 70 91 |

CONCLUSION

So, we can conclude that checking examination papers for patient's 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 **REBERTA** model, we are able to accurately mark the locations of the annotations i.e the symptoms of the patients which have been taken down by the doctors who are taking the exam.

CONCLUSION

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 whether 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.

Thank you



NATURAL LANGUAGE PROCESSING