SIIM CMIMI24
Oct 21-22 | Boston, MA
Conference On Machine Intelligence In Medical Imaging

Utilizing Natural Language Processing and Deep Learning Classification of Radiology Reports to Evaluate the Sensitivity of Chest CT for Detecting Signs of Congestive Heart Failure

Dorina Pinkhasova, Creighton University School of Medicine; Daniel Spalinski; Sherif Zineldine; Ali Memon; Kimberly Mendez, PhD; Michael Fei; Daniel Nguyen; Jad Alsheikh; Randy Richardson, MD

Introduction/Background

There is a lack of considerable data on the use of language models (LMs) in the interpretation of radiology reports, notably chest CTs. Chest CTs in the U.S. have a reported sensitivity and specificity of 86% and 68% in detecting signs of congestive heart failure (CHF).1 LMs can be used to find key characteristics of pathology. These models can allow for enhanced interpretation of reports. This study assesses the capabilities of natural language processing (NLP) in the evaluation of CT reports in patients with a diagnosis of CHF.

Methods/Intervention

This study is a retrospective review of data from the MIMIC-IV, an open-access database derived from the electronic health records of Beth Israel Deaconess Medical Center from 2008 to 2019.2 The multi-label radiology report classification model SARLE was implemented to generate lists of significant findings for chest CTs performed in the same admission with a diagnosis of CHF according to appropriate ICD codes.3 Radiology reads were classified as positive according to the presence of one to six key radiographic findings. An ROC curve was generated based on these varying numbers of positive findings.

Results/Outcome

3,670 hospital admissions where a chest CT was performed with a concurrent diagnosis of CHF were included. Odds ratios (OR, 95% CI) were calculated for each finding using the model interpretation of 91,281 total chest CT reports. These features included cardiomegaly (6.4, 5.9-6.8), vascular congestion (OR 5.8, 4.7-7.2), pleural effusion (OR 8.1, 7.4-8.8), septal thickening (OR 5.2, 4.7-5.7]), pulmonary edema (8.2, 7.6-8.9), and dilated pulmonary vessels (OR 2.3, 2.1-2.5). The presence of at least one key radiographic finding had a 92.34% sensitivity and specificity of 52.92% for the presence of CHF. The resulting ROC curve had an AUC of 0.81 (fig. 1).

Conclusion

NLP allowed for the comprehensive interpretation of a large number of radiographic studies with CHF. Chest CT sensitivity for CHF may be greater than previously reported, and expanding this methodology to other modalities has the potential to better evaluate the accuracy of imaging modalities for specific pathologies.

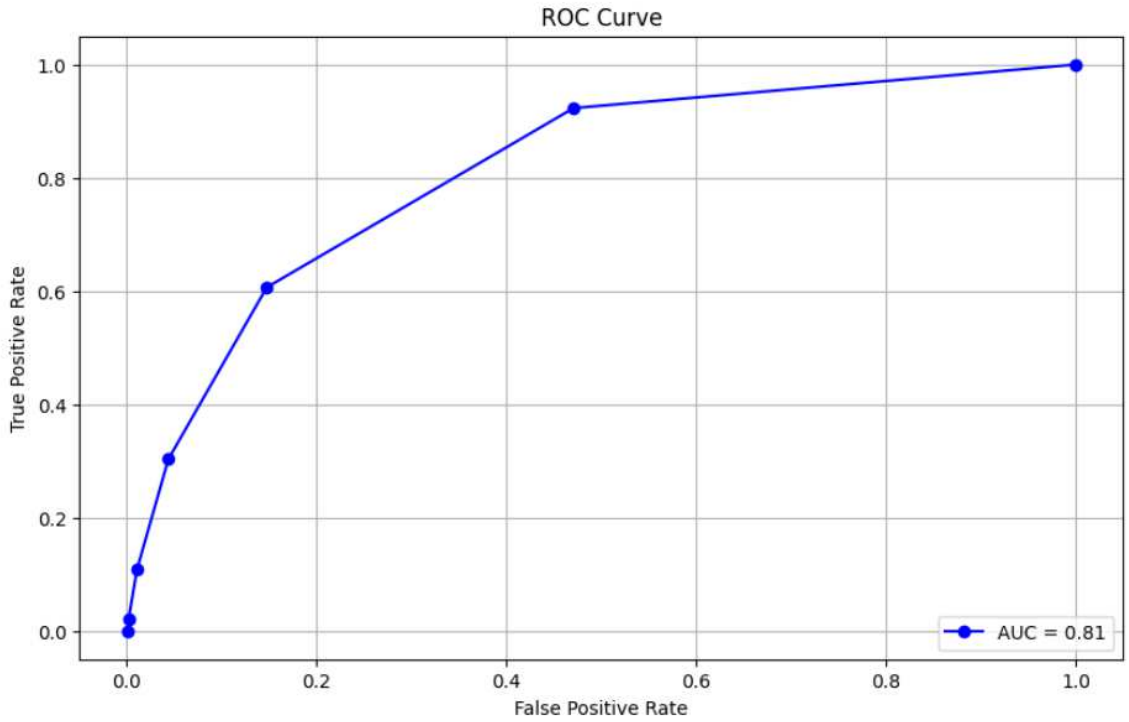
Statement of Impact

The study of NLP has implications for the future of interpretation of radiology reports.

Figure 1.

The figure displays a ROC curve created using Matplotlib with Python to map the use of varying numbers of positive findings on Chest CT to diagnose CHF.

```
Classifying all cases as positive: Sensitivity = 1.0000, Specificity = 0.0000
At least 1 finding(s): Sensitivity = 0.9231, Specificity = 0.5292
At least 2 finding(s): Sensitivity = 0.6070, Specificity = 0.8526
At least 3 finding(s): Sensitivity = 0.3039, Specificity = 0.9563
At least 4 finding(s): Sensitivity = 0.1096, Specificity = 0.9893
At least 5 finding(s): Sensitivity = 0.0226, Specificity = 0.9981
At least 6 finding(s): Sensitivity = 0.0011, Specificity = 0.9999
```



Keywords

Natural Language Processing; Deep Learning; Radiology report classification