Deep learning tool classifies radiology free-text reports

A recent study evaluated the performance of a deep learning convolutional neural network (CNN) model compared with a traditional natural language processing (NLP) model in extracting pulmonary embolism (PE) findings from thoracic computed tomography (CT) reports from two institutions. The findings show that the new CNN model can classify radiology free-text reports with an accuracy of 99%, which is equivalent to or beyond that of an existing traditional NLP model.

"By successfully automating the classification of imaging reports, this work may enable large-scale CNN annotation of free text in medical imaging reports," according to the study led by Stanford University researchers and published in the journal Radiology.

The majority of reports that accompany the imaging examination are composed of free-text narration. This format is the principal reason that most information in medical imaging reports is unstructured and remains inaccessible to automated analysis. Computational NLP methods make it possible to derive useful structured data from large repositories of narrative medical data. Although accurate, NLP techniques require a great deal of development work, including domain-specific feature engineering, complex annotations, and laborious coding for specific tasks.

In contrast, the CNN model used in this study was developed without first defining terms, phrases, or semantic input for the task of identifying PE. The researchers said their CNN model was simply trained on a small proportion of labelled reports with a clearly defined input (impression free text) and output (annotation labels) and achieved superior performance levels.

For the study, contrast material-enhanced CT examinations of the chest performed between January 1998, and January 2016, were selected. Annotations by two human radiologists were made for three categories: the presence, chronicity, and location of PE. Classification of performance of the CNN model with an unsupervised learning algorithm for obtaining vector representations of words was compared with the open-source application PeFinder. Sensitivity, specificity, accuracy, and F1 scores for both the CNN model and PeFinder in the internal and external validation sets were determined.

The CNN model demonstrated an accuracy of 99% and an area under the curve value of 0.97. For internal validation report data, the CNN model had a statistically significant larger F1 score (0.938) than did PeFinder (0.867) when classifying findings as either PE positive or PE negative, but no significant difference in sensitivity, specificity, or accuracy was found. For external validation report data, no statistical difference between the performance of the CNN model and PeFinder was found.

"This highly accurate, generalisable, deep learning-based automated software application for automated classification of radiology free-text reports could be made available for a variety of applications, including diagnostic surveillance, cohort building, quality assessment, labels for computer vision data, and clinical decision support services," the researchers wrote.

Applying the CNN model to reports from a larger range of institutions could help confirm the generalisability of this model and approach, added the researchers.

Source: Radiology
Image Credit: Offutt Air Force Base

Published on : Tue, 27 Feb 2018