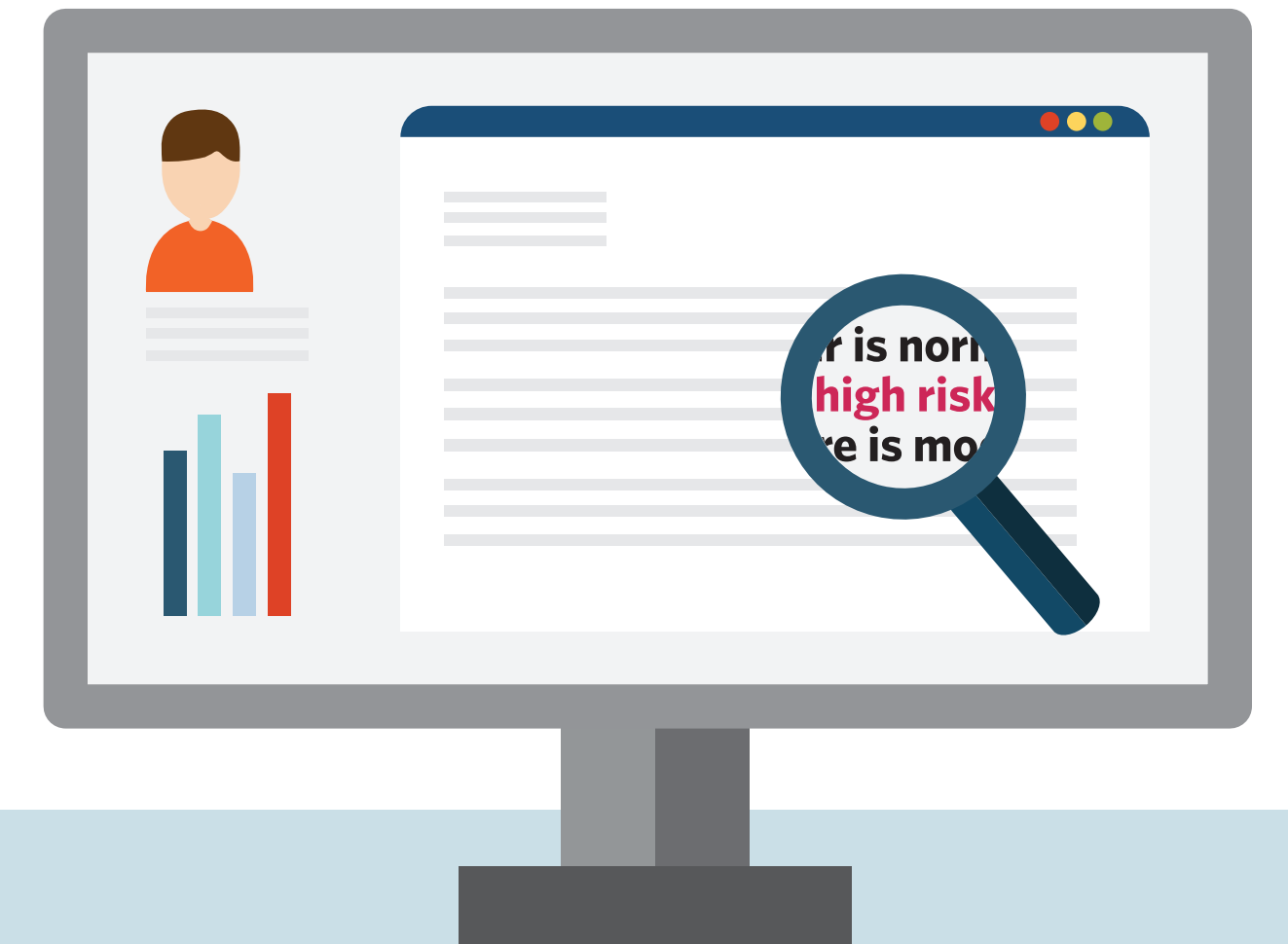


Text-mining of electronic medical records to generate Charlson Comorbidity Index scores



INTRODUCTION

CCI - a prognostic indicator for mortality

When a patient is diagnosed with comorbid conditions, there are two or more diseases existing simultaneously. This occurrence could potentially affect the prognosis of the disease of interest, and subsequently influence its treatment plan.

The Charlson Comorbidity Index (CCI) is commonly used by clinicians to classify the severity of co-existing diseases. It is a "scoring system" that assigns weight to each disease, and

thereafter generates an index of the patient's risk of death. This information would aid the physician in prescribing the best care plan for the patient.

The conventional method of generating a patient's CCI score requires a clinician to first examine comprehensive case notes, and then fill up an extensive questionnaire. These inputs - after calculation - would provide an estimate of the patient's risk of mortality.

As the manual process of perusing each medical record is generally time-consuming, developing a machine-learning algorithm may prove to be a more efficient and systematic alternative.

Objective of study:

- To develop a text-mining algorithm to automate scoring of the CCI by using structured information from electronic databases
- Primary aim of this project is to **improve clinical efficiency and accuracy of diagnoses**, leading to better use of resources and improved patient outcomes

METHODOLOGY



Data source:

- Electronic medical records (both inpatient and SOC) from the Department of Rheumatology & Immunology (RHI) at SGH
- Includes disease diagnosis, clinical notes, prescriptions, treatment and administrative data

Process of developing and validating the algorithm

Two methods of CCI score generation were used in this study:

- Manual calculation by a clinician (these results were regarded as the gold standard)
- Machine-learning algorithm developed to perform text-mining on medical text

Comparison of results



The performance of the algorithm was measured by comparing its results with the clinically verified CCI scores. Predicted scores that were the same or within +/- 1 of the actual scores were treated as accurate.

Results

The algorithm attained an accuracy of:

86.7% for patients whose clinically verified CCI scores were available

Outcomes



The text-mining algorithm would enable clinicians to stratify patients based on similar characteristics, likely risk or diagnosis.



The results may offer insights that could potentially improve prognosis and influence treatment decisions.

Understanding the text-mining algorithm: A form of Named Entity Recognition (NER)

NER is a process of extracting key information within a collection of text and classifying them according to a pre-defined set of categories.

For example, to identify if the patient had myocardial infarction (MI - i.e. heart attack), the algorithm would search clinical records for indications of MI, which would add one point to the CCI.

If explicit indications are found, the algorithm could add this point to the CCI with high confidence.

myocardial infarction

+1



To support the NER process, the machine was trained on medical ontologies and medical references. The UMLS Metathesaurus was used as the base for disease concepts, their synonymous names and their relationships.

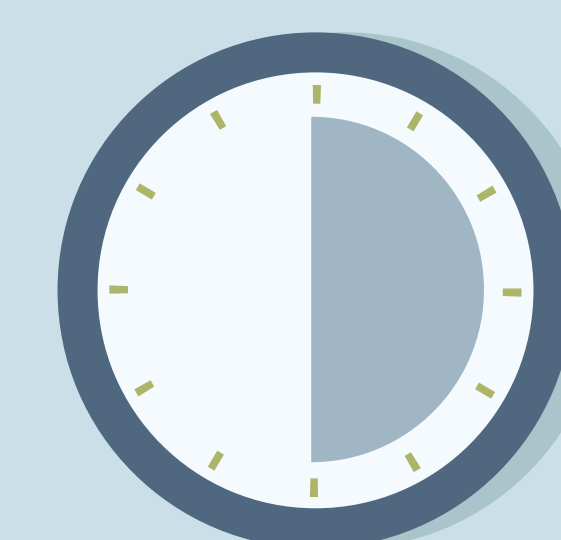
BENEFITS

Saving time and costs

A clinician can save up to an average of 30 minutes if they used the algorithm instead of generating the patient's CCI score manually.

Improved accuracy of clinical diagnosis

While there are merits to both manual and computerised methods of generating CCI scores, the latter would ensure a level of consistency and accuracy with a reliable algorithm. This will help physicians to deliver the most effective treatment plan within a shorter amount of time.



With 30 mins saved per case note, RHI could save up to:

4,778 man hours in one year

\$219,479 over a three-year period

CONCLUSION



This automation could potentially influence treatment decisions and bring about time-savings and efficiency. This text-mining tool could be extended to other departments with high adoption rates of electronic case notes.