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

On-demand Generation of Probabilistic Models for Radiology Differential Diagnosis from Real-world Data

Charles E. Kahn, Jr., MD, MS, FSIIM, University of Pennsylvania

Introduction/Background

Bayesian networks apply probability theory to perform diagnostic reasoning. They offer several attractive features, including the ability to explain their reasoning and to account for missing or conflicting data. However, their construction often is limited by the lack of real-world data to derive the conditional-probability tables (CPTs) that relate two conditions. This work sought to establish an approach to extract probability data from radiology reports and apply that data for on-demand generation of Bayesian network models.

Methods/Intervention

The Radiology Gamuts Ontology (RGO), a reference source of more than 2000 radiology differential-diagnosis listings, was accessed through its application programming interface (https://gamuts.net/api/specialty). Two years of radiology reports from a large U.S. academic health system were analyzed using named-entity recognition and negation-detection techniques to identify positive mentions of RGO entities. An occurrence was defined as positive mention of an RGO entity in a patient. Data were aggregated by patient; the software tallied the number of occurrences of each RGO entity and the number of co-occurrences of each pair of RGO entities. Age and sex distribution of each condition was computed.

Results/Outcome

From approximately 1.8 million reports on 1.3 million distinct patients, the software generated probabilistic data for the 2,742 RGO entities (of 16,839 total) that occurred in the dataset. Project software aggregated probability data around the specified entity and entities that could cause or be caused by it; the generated Bayesian network model was encoded in Structural Modeling, Inference, and Learning Engine (SMILE) format. Diagnostic inference was performed using the GeNIe platform (BayesFusion LLC, Pittsburgh, PA).

Conclusion

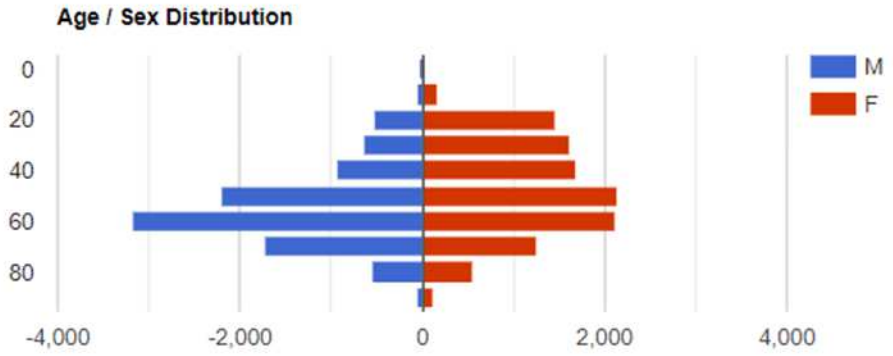
This methodology generates Bayesian-network models for radiology diagnosis from real-world data extracted from analysis of radiology reports. It demonstrates the ability to extract probabilistic data from the unstructured text of radiology reports to generate diagnostic models tailored to the prevalence of diseases and imaging findings of a specific organization's patient population.

Statement of Impact

This report describes a novel approach that generates conditional-probability data from unstructured radiology reports. It overcomes a key limitation of Bayesian networks and allows one to create diagnostic models the apply the frequencies of diseases and imaging findings of a specific set of patients.

ascites

21093 patients
Prevalance: 1.51%



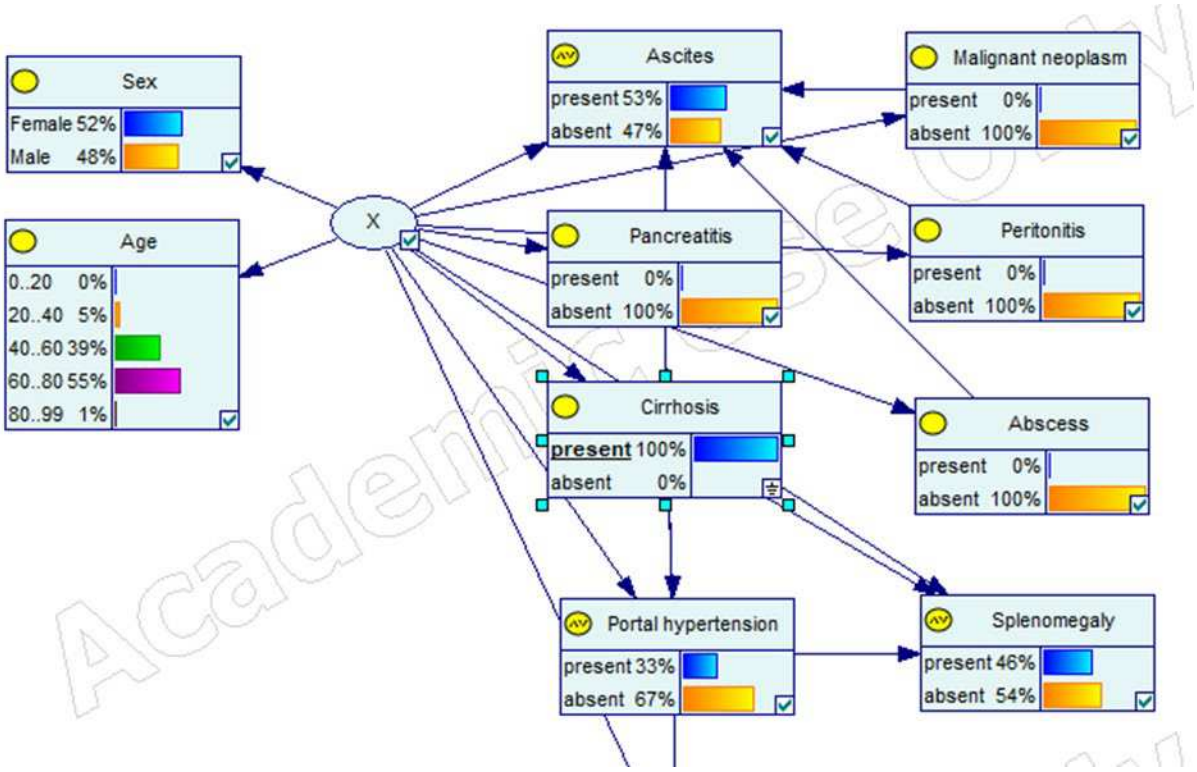
May Cause

pericholecystic fluid (1058)	(667) 3.16 %
abdominal distention (1475)	(660) 3.13 %
peritoneal disease (157)	(118) 0.56 %
cystic pelvic mass (42)	(19) 0.09 %
inferior vena cava obstruction (14)	(6) 0.03 %
middle mediastinal lesion (19)	(4) 0.02 %
small pleural effusion with subsegmental atelectasis (4)	(3) 0.01 %
massive pleural effusion (1)	(1) 0.00 %
bilateral elevated diaphragm (2)	(1) 0.00 %

May Be Caused by

pancreatitis (13534)	(3919) 18.58 %
malignant neoplasm (18645)	(3225) 15.29 %
cirrhosis (5275)	(2738) 12.98 %
peritonitis (5338)	(2581) 12.24 %
abscess (6247)	(1648) 7.81 %

For a specified Radiology Gamuts Ontology entity, here "Ascites," the age and sex distribution is displayed along with conditional probabilities computed based on the entities' co-occurrence.



A portion of the generated Bayesian network model, displayed using the GeNIe platform, that incorporates real-world probabilistic data derived from the corpus of radiology reports.

Keywords

Diagnosis; Bayesian networks; Probabilistic reasoning; Real-world data; Data mining