Journal of the American Medical Informatics Association, 2023, **30(10)**, 1701–1706
https://doi.org/10.1093/jamia/ocad119
Advance access publication 28 June 2023
Research and Applications

Research and Applications

Automated detection of causal relationships among diseases and imaging findings in textual radiology reports

Ronnie A. Sebro[1] and Charles E. Kahn, Jr[2]

[1]Department of Radiology, Department of Orthopedic Surgery, and Center for Augmented Intelligence, Mayo Clinic, Jacksonville, Florida, USA
[2]Department of Radiology and Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, Pennsylvania, USA

*Corresponding Author: Charles E. Kahn, Jr., MD, MS, FACMI, Department of Radiology, University of Pennsylvania, 3400 Spruce Street, Philadelphia, PA 19104, USA; ckahn@upenn.edu

ABSTRACT

Objective: Textual radiology reports contain a wealth of information that may help understand associations among diseases and imaging observations. This study evaluated the ability to detect causal associations among diseases and imaging findings from their co-occurrence in radiology reports.

Materials and Methods: This IRB-approved and HIPAA-compliant study analyzed 1 702 462 consecutive reports of 1 396 293 patients; patient consent was waived. Reports were analyzed for positive mention of 16 839 entities (disorders and imaging findings) of the Radiology Gamuts Ontology (RGO). Entities that occurred in fewer than 25 patients were excluded. A Bayesian network structure-learning algorithm was applied at $P < 0.05$ threshold: edges were evaluated as possible causal relationships. RGO and/or physician consensus served as ground truth.

Results: 2742 of 16 839 RGO entities were included, 53 849 patients (3.9%) had at least one included entity. The algorithm identified 725 pairs of entities as causally related; 634 were confirmed by reference to RGO or physician review (87% precision). As shown by its positive likelihood ratio, the algorithm increased detection of causally associated entities 6876-fold.

Discussion: Causal relationships among diseases and imaging findings can be detected with high precision from textual radiology reports.

Conclusion: This approach finds causal relationships among diseases and imaging findings with high precision from textual radiology reports, despite the fact that causally related entities represent only 0.039% of all pairs of entities. Applying this approach to larger report text corpora may help detect unspecified or heretofore unrecognized associations.

Key words: biomedical ontologies (D064229), data mining (D057225), etiology (Q000209), correlation of data (D000078331), machine learning (D000069550), natural language processing (D009323), radiology (D011871), radiology information systems (D011873)

BACKGROUND AND SIGNIFICANCE

Medical records contain a wealth of information that can be used to better understand diseases and associated conditions. In particular, electronic medical record (EMR) data can be used to discover correlations among diseases and to stratify patient cohorts.[1] Textual information, such as progress notes, imaging reports, and procedural reports, constitutes much of the data available from the EMR; incorporating text-based information can improve the detection and identification of named clinical conditions.[2] In addition, one can use ontologies—a form of knowledge representation that defines the entities in a domain of discourse and their relationships to other entities—to extract clinical information from physicians' free-text notes.[3] Radiology reports have been shown to have a wealth of information about underlying diseases and conditions.[4]

Pairwise co-occurrence statistics can be used to detect associations between diseases and findings.[5,6] A study of co-occurrence data from radiology reports using the phi coefficient (φ) and Cohen's kappa statistic (κ) was able to identify causal relationships at a variety of statistical thresholds.[7] However, that approach showed limited performance, with an area under the receiver operating characteristic curve of approximately 0.7, due in part to the very small proportion of causally related pairs of entities.

A more powerful approach to detect causal relationships is Bayesian network learning. A Bayesian network is a graphical model whose nodes represent random variables, such as diseases and observations.[8,9] Each node has two or more states with associated probability values; for each node, the probability values sum to 1. The directed connections between nodes, called "edges" or "arcs," represent causal or probabilistic influences; each node has a table that expresses the conditional probabilities of the node's states based on those nodes that influence it. Nodes without incoming edges have a probability table that reflects their prior probability, that is, their prevalence. A variety of algorithms can learn the structure of a Bayesian network model directly from data[10,11]; such structure-learning algorithms can account for confounding variables to identify those nodes that are highly likely to be causally linked. The absence of an edge between two nodes indicates that the nodes are conditionally independent. Unlike

Received: 11 December 2022. **Revised:** 10 June 2023. **Editorial Decision:** 13 June 2023. **Accepted:** 16 June 2023

simple co-occurrence statistics, Bayesian networks can identify and account for confounding variables whose influences can cause other variables to falsely appear to be causally related.[12]

OBJECTIVE

The aim of this study was to evaluate the ability of a Bayesian network structure-learning algorithm to detect causal relationships among diseases and imaging findings from their patterns of co-occurrence in textual radiology reports.

MATERIALS AND METHODS

Patient data

The study was approved by the Institutional Review Board and was compliant with the Health Insurance Portability and Accountability Act (HIPAA); the need for signed informed consent from each patient was waived. The study analyzed all reports of radiology examinations performed from January 1, 2015, through December 31, 2016 at a large US health system that serves urban, suburban, and rural populations and has both academic and community-practice settings. Reports were generated as narrative ("free") text or as semistructured text using report templates by more than 190 attending radiologists and more than 100 radiology residents and fellows; final reports were approved by an attending radiologist.

Reports were aggregated by patient to account for imaging findings or disorders that might appear in different examinations, such as the presence of a cerebral artery aneurysm in a patient with autosomal-dominant polycystic kidney disease. Patients' medical record numbers were anonymized and replaced by a unique sequential numerical identifier. The patient's age and gender were recorded; ages greater than 90 years were recorded as 90 years. No protected health information was recorded. The total number of patients and their distribution by age and gender was computed.

Reference ontology

The Radiology Gamuts Ontology (RGO) incorporates a controlled terminology of 16 839 entities that include diseases, interventions, and findings relevant to the practice of radiology.[13] RGO has been integrated with a variety of other biomedical ontologies, including RadLex, SNOMED Clinical Terms, the Disease Ontology, Human Phenotype Ontology, and Orphanet Rare Disease Ontology,[14,15] which allows the incorporation of additional synonyms from those other ontologies. A key aspect of RGO is its representation of 55 619 causal relationships among those entities, such as "*cirrhosis may_cause splenomegaly*." Those causal assertions allow one to formulate differential diagnoses; they also serve as a reference standard ("ground truth") against which one can benchmark the performance of automated systems to detect causal relationships. The *may_cause* relation and its inverse, *may_be_caused_by*, indicate possible causes of an imaging observation. This causal relation is weaker than logical implication and the specified causes of an observation need not be exhaustive.[13] RGO is available through the National Center for Biomedical Ontology's BioPortal website (https://bioportal.bioontology.org/ontologies/GAMUTS).[16]

Identification of entity occurrences

Named entity recognition (NER), a subtype of natural language processing (NLP), was applied to identify the presence in each report of each of RGO's 16 839 disorders, interventions, and imaging findings. In addition to synonyms defined within RGO itself, the search incorporated synonyms from linked ontologies, such as RadLex and SNOMED Clinical Terms. A commercial text-retrieval algorithm with negative-expression filtering was used to identify positive mentions of the terms in the corpus of reports (Nuance mPower, Microsoft, Redmond, WA). We used query expansion to incorporate synonyms and lexical variants: for example, the search terms provided to the mPower algorithm for the Gamuts term "gallstones" included lexical variants such as "gallstone" (from RGO) and "gallbladder stones" (from SNOMED CT) and related terms such as "cholelithiasis" (from RadLex) and "cholecystolithiasis" (from SNOMED CT).

An occurrence was defined as a positive mention of an entity in any of a patient's radiology reports. For example, a patient whose report that indicated "A large right pleural effusion is present" was considered as an occurrence of *pleural effusion*. Negative mentions (eg, "no pneumothorax") and speculative mentions (eg, "rule out AVN") were excluded. To assure appropriate statistical power, entities that occurred in fewer than 25 patients were excluded. The threshold of 25 patients was chosen with the estimate that it would result in 150 RGO entities being included in the analysis, which would result in approximately 10 000 pairwise combinations of entities. Under the assumption that 50% of patients would have the first entity and 4% of those patients would have the second entity, a sample size of 25 patients would have 80% power to detect an association between entities with a Type I error rate of 0.05/10 000.

Data were aggregated by patient. Entities that occurred in reports of the same patient were considered to co-occur, regardless of chronological order of the findings. The cohort for analysis consisted of those patients with at least one of the included entities. Patients who had no positive mention of the included RGO entities were excluded from analysis.

To analyze the performance of the commercial NER system, we sampled two consecutive days of radiology reports from one hospital in the health system: 1277 reports were accrued. For 15 search terms, we evaluated the precision and recall of the NER tool to identify the number of reports with positive mentions (using the system's "Maximize positive findings" setting). Analysis was limited to the specified term; no synonyms or related terms were included. One of the authors (CEK) reviewed the reports to establish ground truth without reference to the automated system's findings. For each search term, precision (TP/(TP+FP)) and recall (TP/(TP+FN)) were computed. The macro-average was computed as the mean of the values for each term; the microaverage was computed from the total number of classified reports; 95% confidence intervals (CI_{95}) were computed for each value.

Detection of causal relationships

We defined a Bayesian network whose nodes represented the included RGO entities; each node had two states, *present* and *absent*. The *bnlearn* hill-climbing algorithm (R version 4.13, R Foundation, Vienna, Austria) was applied to construct a BN model using the Bayesian Information Criterion (BIC) for network scores.[10] No preseeded directed acyclic graph (DAG)

was used to initialize the algorithm, nor was any of RGO's domain knowledge encoded into the BN model. All possible arcs were evaluated for inclusion in the graph. The structure-learning algorithm randomly inserted, removed, or reversed an arc on every iteration; the number of iterations was unlimited. The maximum number of parents for a node was allowed to be one less than the total number of nodes. Score caching was applied to accelerate the structure-learning process.

Statistical analysis was performed using two-sided test statistics and a statistical significance threshold of $P < 0.05$. The data presented to the model consisted of an array with patients by row, RGO entities by column, and a value of 1 or 0 to indicate the presence or absence, respectively, of a positive mention of corresponding entity in any of that patient's reports. The Bayesian network's edges, regardless of direction, were evaluated as possible causal relationships, and were matched against the 55 619 known causal relationships specified in RGO. Those pairs of entities not specified in RGO as causally related were reviewed by two Board-certified radiologists, both with more than 12 years of experience, for plausible causal relationships; decision was achieved by consensus.

The algorithm's precision, recall, positive likelihood ratio (PLR), and negative likelihood ratio (NLR) were computed. Precision, also known as positive predictive value, was computed as the number of true positives—truly causally related pairs of terms, as judged by a matching causal relationship in RGO or by physician review—divided by the total number of pairs identified. Recall (sensitivity) was computed as the number of true positives divided by the number of causal relationships defined in RGO among the included entities. One of the authors (CEK) reviewed a random sample of 250 negative pairs to confirm that they were truly negative (ie, not causally related). PLR is the ratio of true positives to false positives; NLR is the ratio of false negatives to true negatives. Tests with PLR > 10 or NLR < 0.1 are considered useful to establish or exclude an association.[17]

RESULTS

The study cohort included 1 702 462 consecutive reports of 1 396 293 patients; Figure 1 summarizes the distribution of patients by age and sex. Occurrence of entities was recorded for each patient (Figure 2). Of the 16 839 RGO entities, the 2742 entities (16%) that occurred in 25 or more patients were included in the analysis. RGO specified 3804 causal relationships among the included entities. The analysis cohort was comprised of the 53 849 patients (3.9%) who had a positive mention of at least one of the included entities. For each patient, the dataset included age (by decade; ie, values 0, 10, . . ., 90), sex (as assigned at birth), and a Boolean value (present or absent) for each of the 2742 entities.

On the sample of 1277 reports, the NER system achieved a macroaverage precision of 0.992 (CI_{95}, 0.980–1.000) and recall of 0.975 (CI_{95}, 0.956–0.993) (Table 1). Microaverage precision was 0.981 (CI_{95}, 0.960–1.000) and recall was 0.963 (CI_{95}, 0.934–0.992).

The Bayesian network algorithm evaluated more than 12 440 441 possible associations; the final model identified 725 pairs of probabilistically related entities. Of those, 216 (30%) matched a known causal relationship in RGO. Physician review determined that 418 of the 509 (82.1%) other pairs represented plausible causal relationships (Table 2).

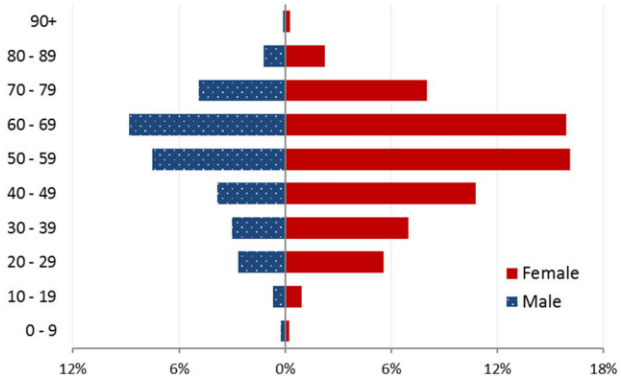

Figure 1. Age and sex distribution of patients in the study cohort.

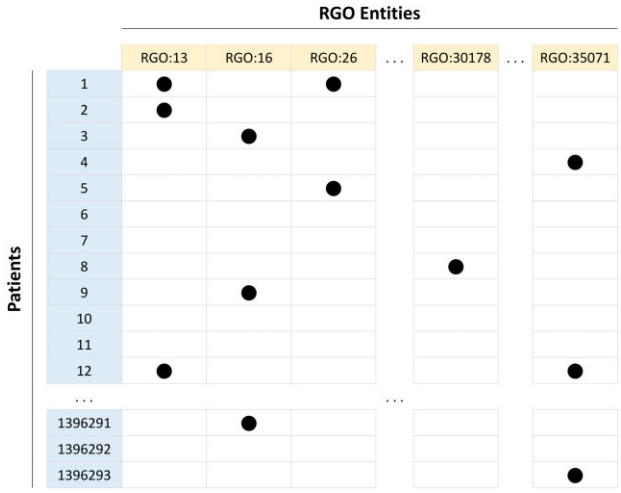

Figure 2. Occurrence of RGO entities in the patient cohort, before excluding rarely occurring entities. The table has 1 396 293 rows (one for each patient) and 16 839 columns (one for each RGO entity). A black dot indicates the occurrence of an RGO entity in that patient.

Table 1. Performance of the commercial named-entity recognition system to detect positive mentions of terms in a sample of 1277 reports

Search term	TP	FP	FN	TN	Precision	Recall
Adenopathy	1	0	0	1276	1.000	1.000
Aortic aneurysm	3	0	0	1274	1.000	1.000
Appendicitis	1	0	0	1276	1.000	1.000
Ascites	13	0	2	1262	1.000	0.867
Cirrhosis	1	0	0	1276	1.000	1.000
Encephalomalacia	7	0	0	1270	1.000	1.000
Hydronephrosis	11	0	3	1263	1.000	0.786
Kidney stone	5	0	0	1272	1.000	1.000
Lymphadenopathy	7	0	0	1270	1.000	1.000
Nephrolithiasis	4	0	0	1273	1.000	1.000
Parkinson's disease	2	0	0	1275	1.000	1.000
Pleural effusion	56	2	0	1219	0.966	1.000
Pneumonia	28	0	1	1248	1.000	0.966
Pneumothorax	10	1	0	1266	0.909	1.000
Splenomegaly	7	0	0	1270	1.000	1.000
Total	156	3	6	18 990		
				Macroaverage	0.992	0.975
				Microaverage	0.981	0.963

The "true positive" (TP) count was the number of reports that correctly identified a positive mention of the search term. False positive (FP), false negative (FN), and true negative (TN) counts are tallied.

Table 2. Six examples of entity pairs identified by the Bayesian network

	Causally related		
Entity pair	RGO	Physician review	Conclusion
Leukemia—thrombocytopenia	Yes	—	Yes
Gastroenteritis—vomiting	Yes	—	Yes
Bladder cancer—ileal conduit	No	Yes	Yes
Desmoid tumor—mesenteric mass	No	Yes	Yes
Meningioma—paranasal sinusitis	No	No	No
Conjunctivitis—ligamentum flavum thickening	No	No	No

The "RGO" column indicates whether the Radiology Gamuts Ontology asserted a *may_cause* relationship between the specified entities. The "Physician Review" column indicates if physician review determined that the entities have a plausible causal relationship. The rightmost column summarizes the two determinations.

Table 3. Results of Bayesian network model for the 2742 included entities that occurred in 25 or more patients

		Number of pairs of entities		
		Related	Unrelated	Total
Causal relationship identified by Bayesian network algorithm	Yes	634	91	725
	No	3170	3 754 016	3 757 186
	Total	3804	3 754 107	3 757 911

The total number of possible entity pairs is computed as $n\cdot(n-1)/2$, where $n=2742$.

Thus, the Bayesian network learning algorithm achieved precision (positive predictive value) of 87% (634/725) (Table 3). Of the 3804 causal relationships among the 2742 included entities, the 216 RGO causally related pairs found by the algorithm yielded a recall (sensitivity) of 5.7% (216/3804). However, the PLR of 6876 indicates that the algorithm substantially increased the odds of detecting a true causal association.

The NLR of 0.83 indicates that the algorithm had limited ability to discern the absence of a causal relation; however, that limitation was not of concern given that 99.96% of all possible entity pairs had no causal relationship defined in RGO. Manual review of 250 entity pairs selected randomly from the 141 766 816 negative pairs (141 767 541 possible entity pairs, less the 725 pairs found to be causally related) found none to have a causal relationship; hence, the specificity of the algorithm approximated 1 (CI_{95}, 0.998–1).

DISCUSSION

We applied a Bayesian network structure-learning algorithm to identify causal relationships from data indicating the occurrence of diseases and imaging findings in a large set of textual radiology reports. The algorithm yielded a model with 725 edges, each of which specified a pair of probabilistically related entities from RGO. Of those 725 pairs, 634 were judged to be "true positives" to yield an overall precision of 87%. Most of the related entities identified by the algorithm were not specified in the RGO, which suggests that the technique applied here could help discover heretofore unspecified causal associations. The PLR of 6876 indicates that the algorithm very effectively improved detection of causal associations.

The 5.2% recall reflects the extremely low frequency of causally related entities: RGO's 55 619 known causal relationships constituted less than 0.04% of the 141 767 541 possible pairs of 16 839 entities. Additional data should allow a more powerful analysis and investigation is now underway using an order of magnitude more reports. Although the recall (sensitivity) remains low, the Bayesian-network algorithm shows promise to identify causal associations between entities.

This study had several limitations. Many of RGO's disorders and imaging findings are quite rare, and were observed in few, if any, of the 1.3 million patients. The study did not account for supertype–subtype ("subsumption") relationships among RGO entities. Because patients did not have every type of imaging study, related conditions may not have been observed. For example, there are known associations between sickle cell disease (SCD) and H-shaped thoracolumbar vertebrae, and between SCD and frontal bossing of the skull; a patient with SCD who had thoracolumbar radiographs but not skull radiographs would contribute to detecting the former, but not the latter. All reports came from a single organization, albeit one with diverse settings and populations; reports were dictated by more than 190 attending radiologists and 100 trainees. The current analysis did not account for temporal relationships among terms; although the algorithm may have identified a causal relationship between trauma and fracture, for example, there was no information that one followed the other. Several other methods could be used to extract causal relationships from data, including causal mediation analysis,[18] Mendelian randomization,[19] propensity score matching analysis,[20] and Bayesian networks using the PC algorithm.[21]

The commercial text-search algorithm showed a strong ability to detect positive mentions of entities in the current text corpus. The algorithm's negative-expression ("negex") filtering techniques have identified positive and negative instances of imaging findings with accuracy in excess of 97%.[22] PRESTO, the predecessor to the commercial system used in the current study, showed 91% precision and 97% recall in detecting adrenal findings in 32 974 abdominal CT reports.[23] Our audit of a sample of 1277 reports from the current study estimated the macroaverage precision and recall of 0.992 and 0.975, respectively.

At least 90% of the reports in the study cohort were based on reporting templates that fixed the order of elements and in many cases, provided limited "pick-list" options for the values shown in the reports. In general, though, the labels for the various fields of the reports did not specify a diagnosis, but rather an anatomical site. For example, for the "Gallbladder" field, the "normal" option indicated "No calcified gallstones. No gallbladder wall thickening." Here, negation of the specific findings ("gallstones" and "gallbladder wall thickening") precluded that the system would identify them as positive occurrences. Although no report was observed in which a label or heading was counted as an occurrence, we did not conduct a systematic evaluation of such a possibility. Because we defined an "occurrence" at the level of a patient, the presence of repeated positive mentions of a term in a single report—for example, in both the narrative ("findings") section and impression section—did not result is false positives.

In most settings, the great majority of information within an EHR is in the form of narrative text. NLP has been applied EHR systems to support a variety of clinical applications such as representation learning, information extraction, outcome prediction, and phenotyping.[24,25] Radiology reports comprise a large and important subset of textual EHR data and can form a valuable source of data for deep phenotyping.[26,27] NLP systems have shown high performance in identifying and classifying radiology reports, and offer a promising approach to extract measurable information from conventional narrative ("free-text") reports.[28,29] An NLP system successfully determined the presence of more than 20 clinical indications and imaging findings from a database of 889 921 chest radiographic reports.[30] Text mining of radiology reports has shown a strong ability to identify limb fractures, pneumonia, and hepatocellular carcinoma stage.[31–34] Rare diseases have shown an approximate, but significant correlation between their known prevalence and their frequency of occurrence in radiology reports.[35] The availability of radiology report data and tools to mine those data allow one to establish causal rules.[36]

A number of statistical and computational approaches have been developed to infer causal relationships from data.[37] Bayesian networks provide a graph-based causal inference model. A Bayesian network is a DAG where each variable is represented as a node, and a connection between nodes is represented as a directed edge with associated conditional probability table. Variables that are independent remain unconnected. Bayesian networks require three assumptions to ensure valid causal inference: (1) faithfulness—that is the dependence relationships in the Bayesian network hold in the data; (2) causal sufficiency—where all common causes of pairs in the set of variables are present in the analysis; and (3) the causal Markov condition (CMC)—that each node is independent of its nondescendants, conditioned on its parents (direct causes). Bayesian network analysis can be susceptible to selection bias: if missing data are not completely random (independent of all variables), the Bayesian network's inference may be invalid. However, the high precision of the Bayesian network's selection of causally related entities suggests that any missing data were sufficiently random.

The current study falls broadly under the category of causal machine learning, in which one seeks not only to identify patterns of co-occurrence or correlation, but to achieve a deeper understanding of the causal relationships among diseases and their imaging phenotypes.[38,39] In medical imaging, causal reasoning can help address major challenges in machine learning, such as the paucity high-quality annotated data and mismatches between training data and the setting in which the machine learning system will be used.[40] Causal knowledge also allows one to assess a system would react to an intervention and to make robust, actionable decisions in the presence of confounding information.[41]

CONCLUSION

A Bayesian network structure-learning algorithm identified causally related diseases and imaging findings with high precision. The algorithm identified numerous relationships among diseases and imaging findings that were plausible, but previously unspecified in an ontology of radiological differential diagnosis. Given the breadth of topics covered by radiology reports, such analyses offer powerful approaches to identify causally related disorders and imaging phenotypes. Simple pairwise co-occurrence metrics achieve poor performance, for a variety of reasons, including the potentially confounding effects of co-occurring conditions. Further investigation is underway to apply these methods to substantially larger corpora of radiology reports. Analysis of imaging reports holds promise to validate known associations and to identify new associations among diseases and imaging findings.

AUTHOR CONTRIBUTIONS

Both authors designed the experiment. CEK collected the data. RAS performed the analysis. Both authors contributed to writing and reviewing the manuscript.

CONFLICT OF INTEREST

The authors have no competing interest to declare.

DATA AVAILABILITY

The anonymized dataset will be made available upon publication.

REFERENCES

1. Roque FS, Jensen PB, Schmock H, *et al.* Using electronic patient records to discover disease correlations and stratify patient cohorts. *PLoS Comput Biol* 2011; 7 (8): e1002141. doi: 10.1371/journal.pcbi.1002141.
2. Ford E, Carroll JA, Smith HE, Scott D, Cassell JA. Extracting information from the text of electronic medical records to improve case detection: a systematic review. *J Am Med Inform Assoc* 2016; 23 (5): 1007–15. doi: 10.1093/jamia/ocv180.
3. Yehia E, Boshnak H, AbdelGaber S, Abdo A, Elzanfaly DS. Ontology-based clinical information extraction from physician's free-text notes. *J Biomed Inform* 2019; 98: 103276. doi: 10.1016/j.jbi.2019.103276.
4. Hassanpour S, Langlotz CP. Information extraction from multi-institutional radiology reports. *Artif Intell Med* 2016; 66: 29–39. doi: 10.1016/j.artmed.2015.09.007.
5. Cao H, Hripcsak G, Markatou M. A statistical methodology for analyzing co-occurrence data from a large sample. *J Biomed Inform* 2007; 40 (3): 343–52. doi: 10.1016/j.jbi.2006.11.003.
6. Cao H, Markatou M, Melton GB, Chiang MF, Hripcsak G. Mining a clinical data warehouse to discover disease-finding associations using co-occurrence statistics. *AMIA Annu Symp Proc* 2005; 2005: 106–10.
7. Sebro R, Kahn CE Jr. Causal associations among diseases and imaging findings in radiology reports. *Stud Health Technol Inform* 2022; 294: 411–2. doi: 10.3233/shti220487.
8. Pearl J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann Publishers; 1988.
9. Jensen FV. *An Introduction to Bayesian Networks*. New York: Springer; 1996.
10. Scutari M. Learning Bayesian networks with the bnlearn R package. *J Stat Soft* 2010; 35 (3): 1–22. doi: 10.18637/jss.v035.i03.
11. Tsamardinos I, Brown LE, Aliferis CF. The max-min hill-climbing Bayesian network structure learning algorithm. *Mach Learn* 2006; 65 (1): 31–78.
12. Cooper GF. *An Overview of the Representation and Discovery of Causal Relationships Using Bayesian Networks. Computation, Causation, and Discovery*. Menlo Park, CA: AAAI Press; 1999: 4–62.

13. Budovec JJ, Lam CA, Kahn CE Jr. Radiology Gamuts Ontology: differential diagnosis for the Semantic Web. *Radiographics* 2014; 34 (1): 254–64. doi: 10.1148/rg.341135036.
14. Filice RW, Kahn CE Jr. Integrating an ontology of radiology differential diagnosis with ICD-10-CM, RadLex, and SNOMED CT. *J Digit Imaging* 2019; 32 (2): 206–10. doi: 10.1007/s10278-019-00186-3.
15. Kahn CE Jr. Integrating ontologies of rare diseases and radiological diagnosis. *J Am Med Inform Assoc* 2015; 22 (6): 1164–8. doi: 10.1093/jamia/ocv020.
16. Noy NF, Shah NH, Whetzel PL, *et al.* BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Res* 2009; 37 (Web Server issue): W170–3. doi: 10.1093/nar/gkp440.
17. Deeks JJ, Altman DG. Diagnostic tests 4: Likelihood ratios. *BMJ* 2004; 329 (7458): 168–9. doi: 10.1136/bmj.329.7458.168.
18. Zhang Z, Zheng C, Kim C, Poucke SV, Lin S, Lan P. Causal mediation analysis in the context of clinical research. *Ann Transl Med* 2016; 4 (21): 425. doi: 10.21037/atm.2016.11.11.
19. Sanderson E, Glymour MM, Holmes MV, *et al.* Mendelian randomization. *Nat Rev Methods Primers* 2022; 2 (1): 6. doi: 10.1038/s43586-021-00092-5.
20. Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med* 2008; 27 (12): 2037–49. doi: 10.1002/sim.3150.
21. Tsagris M. Bayesian network learning with the PC algorithm: an improved and correct variation. *Appl Artif Intell* 2019; 33 (2): 101–23. doi: 10.1080/08839514.2018.1526760.
22. Dang PA, Kalra MK, Blake MA, *et al.* Use of Radcube for extraction of finding trends in a large radiology practice. *J Digit Imaging* 2009; 22 (6): 629–40. doi: 10.1007/s10278-008-9128-x.
23. Zopf JJ, Langer JM, Boonn WW, Kim W, Zafar HM. Development of automated detection of radiology reports citing adrenal findings. *J Digit Imaging* 2012; 25 (1): 43–9. doi: 10.1007/s10278-011-9425-7.
24. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J Biomed Health Inform* 2018; 22 (5): 1589–604. doi: 10.1109/jbhi.2017.2767063.
25. Zeng Z, Deng Y, Li X, Naumann T, Luo Y. Natural language processing for EHR-based computational phenotyping. *IEEE/ACM Trans Comput Biol Bioinform* 2019; 16 (1): 139–53. doi: 10.1109/TCBB.2018.2849968.
26. Robinson PN. Deep phenotyping for precision medicine. *Hum Mutat* 2012; 33 (5): 777–80. doi: 10.1002/humu.22080.
27. Son JH, Xie G, Yuan C, *et al.* Deep phenotyping on electronic health records facilitates genetic diagnosis by clinical exomes. *Am J Hum Genet* 2018; 103 (1): 58–73. doi: 10.1016/j.ajhg.2018.05.010.
28. Cai T, Giannopoulos AA, Yu S, *et al.* Natural language processing technologies in radiology research and clinical applications. *Radiographics* 2016; 36 (1): 176–91. doi: 10.1148/rg.2016150080.
29. Linna N, Kahn CE Jr. Applications of natural language processing in radiology: a systematic review. *Int J Med Inform* 2022; 163: 104779. doi: 10.1016/j.ijmedinf.2022.104779.
30. Hripcsak G, Austin JH, Alderson PO, Friedman C. Use of natural language processing to translate clinical information from a database of 889,921 chest radiographic reports. *Radiology* 2002; 224 (1): 157–63. doi: 10.1148/radiol.2241011118.
31. Asatryan A, Benoit S, Ma H, English R, Elkin P, Tokars J. Detection of pneumonia using free-text radiology reports in the BioSense system. *Int J Med Inform* 2011; 80 (1): 67–73. doi: 10.1016/j.ijmedinf.2010.10.013.
32. Wagholikar A, Zuccon G, Nguyen A, *et al.* Automated classification of limb fractures from free-text radiology reports using a clinician-informed gazetteer methodology. *Australas Med J* 2013; 6 (5): 301–7. doi: 10.4066/AMJ.2013.1651.
33. Yim WW, Kwan SW, Johnson G, Yetisgen M. Classification of hepatocellular carcinoma stages from free-text clinical and radiology reports. *AMIA Annu Symp Proc* 2017; 2017: 1858–67.
34. Chen MC, Ball RL, Yang L, *et al.* Deep learning to classify radiology free-text reports. *Radiology* 2018; 286 (3): 845–52. doi: 10.1148/radiol.2017171115.
35. Kahn CE Jr. An ontology-based approach to estimate the frequency of rare diseases in narrative-text radiology reports. *Stud Health Technol Inform* 2017; 245: 896–900.
36. Jin Z, Li J, Liu L, Le TD, Sun B, Wang R. Discovery of causal rules using partial association. In: *IEEE 12th International Conference on Data Mining*. IEEE: Brussels, Belgium; 2012: 309–18. doi: 10.1109/ICDM.2012.36.
37. Kleinberg S, Hripcsak G. A review of causal inference for biomedical informatics. *J Biomed Inform* 2011; 44 (6): 1102–12. doi: 10.1016/j.jbi.2011.07.001.
38. Malinsky D, Danks D. Causal discovery algorithms: a practical guide. *Philos Compass* 2018; 13 (1): e12470. doi: 10.1111/phc3.12470.
39. Schölkopf B, Locatello F, Bauer S, *et al.* Toward causal representation learning. *Proc IEEE* 2021; 109 (5): 612–34.
40. Castro DC, Walker I, Glocker B. Causality matters in medical imaging. *Nat Commun* 2020; 11 (1): 3673. doi: 10.1038/s41467-020-17478-w.
41. Sanchez P, Voisey JP, Xia T, Watson HI, O'Neil AQ, Tsaftaris SA. Causal machine learning for healthcare and precision medicine. *R Soc Open Sci* 2022; 9 (8): 220638. doi: 10.1098/rsos.220638.