cmajOPEN

Research

Hypertension identification using inpatient clinical notes from electronic medical records: an explainable, data-driven algorithm study

Elliot A. Martin PhD, Adam G. D'Souza PhD, Seungwon Lee PhD MPH, Chelsea Doktorchik MSc, Cathy A. Eastwood PhD, Hude Quan MD PhD

Abstract

Background: Case identification is important for health services research, measuring health system performance and risk adjustment, but existing methods based on manual chart review or diagnosis codes can be expensive, time consuming or of limited validity. We aimed to develop a hypertension case definition in electronic medical records (EMRs) for inpatient clinical notes using machine learning.

Methods: A cohort of patients 18 years of age or older who were discharged from 1 of 3 Calgary acute care facilities (1 academic hospital and 2 community hospitals) between Jan. 1 and June 30, 2015, were randomly selected, and we compared the performance of EMR phenotype algorithms developed using machine learning with an algorithm based on the Canadian version of the *International Statistical Classification of Diseases and Related Health Problems*, *10th Revision* (ICD), in identifying patients with hypertension. Hypertension status was determined by chart review, the machine-learning algorithms used EMR notes and the ICD algorithm used the Discharge Abstract Database (Canadian Institute for Health Information).

Results: Of our study sample (n = 3040), 1475 (48.5%) patients had hypertension. The group with hypertension was older (median age of 71.0 yr v. 52.5 yr for those patients without hypertension) and had fewer females (710 [48.2%] v. 764 [52.3%]). Our final EMR-based models had higher sensitivity than the ICD algorithm (> 90% v. 47%), while maintaining high positive predictive values (> 90% v. 97%).

Interpretation: We found that hypertension tends to have clear documentation in EMRs and is well classified by concept search on free text. Machine learning can provide insights into how and where conditions are documented in EMRs and suggest nonmachine-learning phenotypes to implement.

Condition identification is an essential part of a learning health system,[1] monitoring health system performance and risk adjustment. The gold standard for case identification is typically chart review.[2] This requires a substantial time commitment from professionals — often making it infeasible for population research. Coded administrative data are commonly used to identify conditions but often have low sensitivity.[3] Electronic medical record (EMR) phenotypes can be automated, making them relatively inexpensive to implement, and have the potential to have both high sensitivity and positive predictive value (PPV).[2] Data from EMRs could supplement administrative data for health research, which could assist in clinical decision-making processes.

Administrative health databases have been used for hypertension surveillance because the data are routinely collected, cover large geographic areas and have the potential for longitudinal follow-up.[4,5] A hypertension case definition was developed using administrative data coded by the *International Classification of Diseases* (ICD), with a reported sensitivity of 68.3%, a PPV of 93.1%, a specificity of 97.8% and a negative predictive value (NPV) of 87.7%.[3,6] The observed undercoding of hypertension was possibly due to the process of coding health information to administrative data. Specifically, the undercoding could be attributed to coders having limited time with a quota of 25 charts per day in Alberta and chart incompleteness, with an estimated 80% of charts missing the discharge summary or operative report at the time of coding.[7,8] Extracting collected health information from EMRs is a promising opportunity to improve the accuracy of identifying hypertension. Clinical notes are a rich source of information in EMRs but are underused in automated processes owing to the difficulties in extracting information from free text. The Unified Medical Language System (UMLS)[9]

Competing interests: None declared.

This article has been peer reviewed.

Correspondence to: Elliot Martin, eamartin@ucalgary.ca

CMAJ Open **2023 February 14. DOI:10.9778/cmajo.20210170**

attempts to overcome some of these difficulties by mapping the varying lexical choices available in clinical documentation to a single concept unique identifier (CUI). We hypothesized that CUIs could play an important role in creating interpretable models. Our objective was to develop a standardized hypertension case identification method using inpatient clinical notes from EMRs. This work is part of our larger research program on EMR phenotyping.[2,10–12]

Methods

Study design

We conducted a retrospective cohort study in which we reviewed medical charts to determine a reference standard for hypertension status in a cohort of randomly selected patients who were admitted to 3 acute care centres in Calgary, Alberta.[13] We compared the performance of EMR phenotype algorithms developed using machine learning and an ICD-based algorithm in identifying patients with hypertension. We used machine learning to develop an EMR phenotype for hypertension and create a data-driven, rule-based algorithm. As a comparative method, we also employed a previously validated hypertension case-identification algorithm based on the Canadian version of the *International Statistical Classification of Diseases and Related Health Problems, 10th Revision* (ICD-10-CA) (termed ICD algorithm in this paper). We compared the performance of both phenotypes against the chart review as a reference standard. We reported our study results using the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) checklist.[14]

Study setting and participants

We used data from a chart review cohort of patients who were selected for a field trial of the new *International Statistical Classification of Diseases and Related Health Problems, 11th Revision for Mortality and Morbidity Statistics* (ICD-11) diagnosis coding system.[13] The patients in this cohort were at least 18 years of age and were discharged from 1 of 3 acute care facilities in Calgary, between Jan. 1 and June 30, 2015. The 3 facilities involved were an academic hospital (Foothills Medical Centre, affiliated with the University of Calgary) and 2 community hospitals (Peter Lougheed Centre and Rockyview General Hospital). These were chosen because they were the inpatient facilities in Calgary with emergency departments and intensive care units that had the same EMR system at that time.

We excluded obstetric admissions owing to short stays and few chronic conditions. We randomly selected patient discharges from this period such that each hospital had about the same representation. Each patient had the same probability of being selected as any other qualifying patient from the same hospital. For patients with several admissions, 1 admission within the study period was randomly selected. A sample-size calculation from the ICD-11 field trial study[13] determined that 3000 records were required to detect a 10% difference in sensitivity between ICD-10-CA and ICD-11 coding of common comorbidities, such as hypertension, using Lachenbruch's midpoint method[15] based on prevalence results from previous findings.[3]

Data sources

We used 3 databases to conduct this study: Sunrise Clinical Manager (SCM), the Alberta Discharge Abstract Database (DAD) and a medical chart review database, linked using a personal health number (a unique lifetime identifier), chart number (a unique number associated with a patient's admission) and admission date.

Sunrise Clinical Manager

This database supplied the clinical notes that were used to extract features to train our machine-learning models. AllScripts SCM is a city-wide, population-level EMR system currently in operation throughout all acute care facilities in Calgary. Alberta Health Services, the single health authority in Alberta, manages SCM and the associated electronic data warehouse.[10] All physician and nursing notes, excluding diagnostic imaging reports and those associated with patients' visits, in our cohort were extracted. We also extracted the name of the clinical notes and identified 58 unique names (e.g., Discharge Summary — Medical, Surgical Assessment and History — Nursing, History and Physical, and others) in our cohort. They included physician and nursing notes but not diagnostic imaging reports.

Discharge Abstract Database

This database supplied the ICD codes for the validated ICD hypertension algorithm as well as age, sex and physician specialty, which we stratified by. The DAD is the administrative health database where up to 25 diagnosis codes for all inpatient encounters are stored using ICD-10-CA.[6] The diagnosis codes are assigned by coders after discharge, based on the clinical documentation in patients' charts. The database also contains basic demographic information about the patients (e.g., sex and age). The Canadian Institute for Health Information provides national coding standards and training programs for health information managers (i.e., coders).[16] It is common practice to identify patients with hypertension who were admitted to hospital using a case definition based on ICD-10-CA codes from any of the 25 diagnosis fields within DAD (I10–I13 or I15).[6] We extracted physician specialty groups to allow grouping of patients under care to surgical and medical groups. We stratified by surgical versus nonsurgical services because preliminary analysis showed many patients were missing discharge summaries in their EMR, and these were predominantly surgical patients. We classified all DAD records with a column listing the specialty of the most responsible physician (DOCSVC1) that corresponded to a specialty of general surgery, orthopedic surgery, vascular surgery, neurosurgery, plastic surgery, thoracic surgery, cardiac surgery and oral surgery as surgical patients.

Medical chart review

This provided the reference standard for hypertension that both our algorithm and the existing ICD algorithm were compared against. We extracted patient charts for each of the included admissions from the hospital records departments,[13] which provided EMR data from SCM as well as

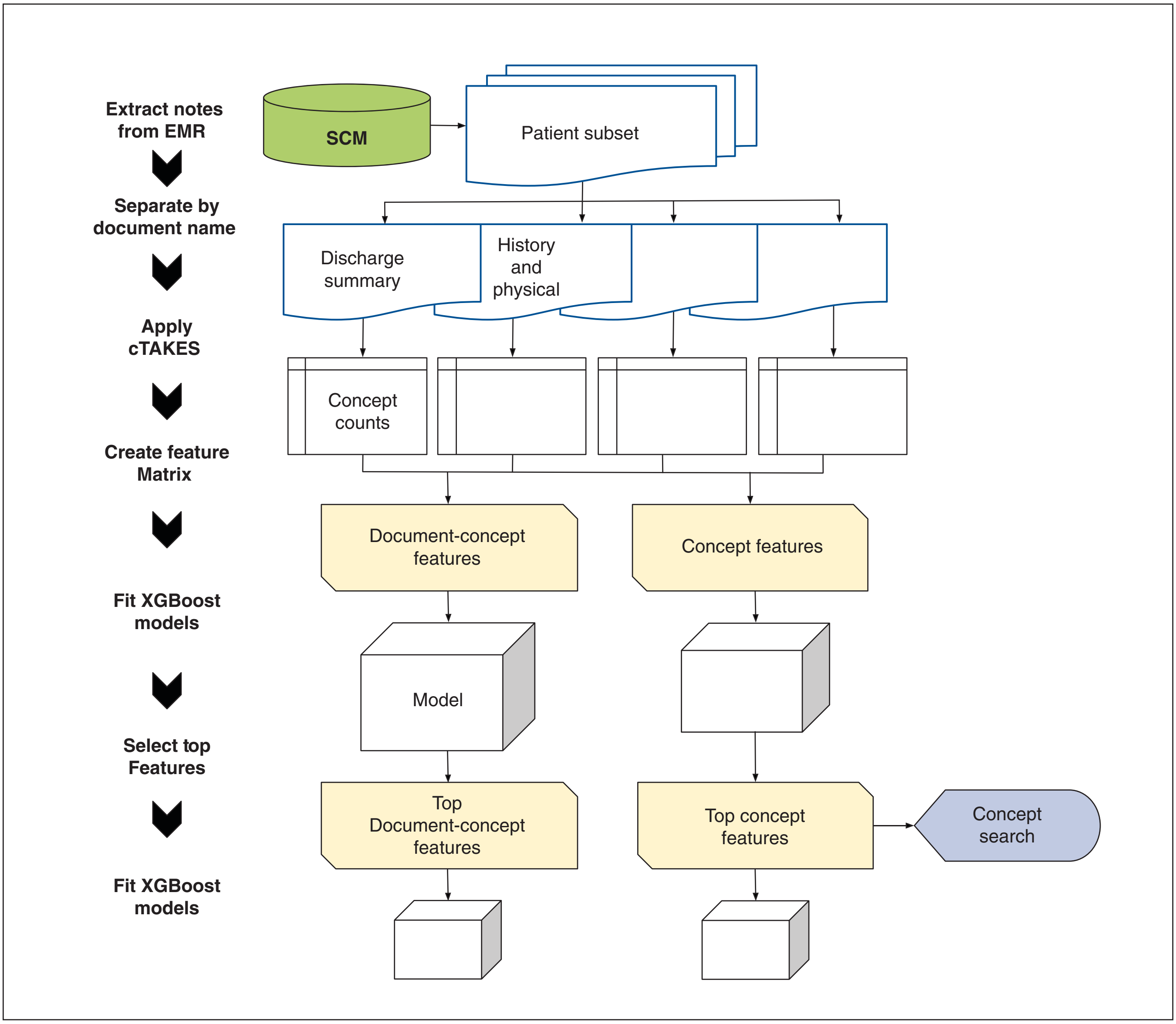

Figure 1: Case identification flow chart. We randomly sampled 3040 inpatient charts from Sunrise Clinical Manager (SCM) and extracted their associated clinical notes, identifying UMLS concepts with cTAKES. We used XGBoost models to select the most important concept and document–concept pair features separately. We used these selected features to fit reduced concept and document–concept XGBoost models. We also used the concept features to implement a simple search algorithm for the hypertension concept C0020538. Note: cTAKES = clinical Text Analysis and Knowledge Extraction System, EMR = electronic medical record, UMLS = Unified Medical Language System.

paper charts not necessarily a part of the EMR. Trained chart reviewers (all nurses) looked for any form of listed diagnosis of hypertension (e.g., controlled hypertensives or essential hypertension) in patients' history and physical, multidisciplinary progress notes, consult notes and discharge summary. If a diagnosis was documented, the chart was labelled as hypertension present. The chart reviewers were not trying to ascertain if patients met diagnostic criteria for hypertension, but only if they had a documented diagnosis of hypertension. The inter-rater reliability between reviewers was high (> 0.8 κ).[13] These hypertension labels were used as a reference standard for both our machine-learning algorithms and for the ICD-based algorithm.

Development of the case definition using machine learning

We outline the steps from extracting the EMR data to our final hypertension case-identification algorithms in Figure 1. The data were split into 80% training and 20% test, which is common in machine learning.[17,18] The training set was used for training and validation of all the machine-learning models, via fivefold cross-validation, and the test set was used only to compare the performance of the final EMR models with the ICD method.

Applying cTAKES and creating a feature matrix

We used the clinical Text Analysis and Knowledge Extraction System (cTAKES),[19] in particular its default clinical pipeline, to process all the clinical notes. We extracted clinical concepts in the form of CUIs from the UMLS.[9] This method accounts for variation in terminology among EMRs, because UMLS maps synonymous terms to the same underlying concept. For example, in UMLS, the clinical concept "hypertensive disease" is assigned the CUI "C0020538." The 2018AB UMLS release contains 67 synonyms for this clinical concept, including "BLOOD PRESSURE HIGH," "HBP," "HTN," "hyperpiesia," "hypertension" and "systemic HTN."

All of these synonyms map to the same CUI, which allowed us to generate nonredundant (i.e., normalized) features. We used the negation and subject attribute annotators in cTAKES to label each CUI. These assessed whether the concept appeared in a negated context (e.g., "no evidence of hypertension") and whether the subject to which the CUI was associated was the patient or someone else. The cTAKES outputs were then converted into a document–concept matrix containing the counts of each CUI for each differently named document ("document") and each chart. Only CUIs that had the patient as their subject and that cTAKES determined were non-negated were counted.

Fitting machine learning models and selecting top features

Feature selection is the process of identifying the variables most relevant to the problem. Our features included both the CUIs and the clinical notes that could discriminate cases of hypertension. There were 58 unique clinical note names in our extracted EMR data, such as "discharge summary" and "history and physical." We used these to create 2 different types of feature sets. The first set of concept features contained only the number of times each concept occurred for each patient; the second set of document–concept features contained the number of times each concept appeared in documents with a given name. For example, the counts of history_and_physical-C0020538 and discharge_summary_medical-C0020538 would contribute to the same C0020538 feature in the first set and would be separate features in the second set. The first set of features could illustrate the most reliable concepts used to identify hypertension, whereas the second could illustrate the most high-yield and trustworthy documents to look at for future chart review.

We estimated the relative importance of each feature for determining hypertension using the gradient boosted algorithm XGBoost, commonly used for supervised learning problems.[20] For each feature set, 5 XGBoost models were fit, each using fivefold cross validation[21] optimizing for area under the receiver operating characteristic curve (AUC) (Appendix 1, Supplementary Table 1, available at www.cmajopen.ca/content/11/1/E131/suppl/DC1, for grid search parameters). The training data were split into 5 groups with about the same number of visits and proportion of patients with hypertension. Five models were then trained, each using a different fold (group) as the validation set to test performance and trained on the remaining 4 folds. This was done to ensure that only reliable features were selected, and to exclude those that only performed well on a subset of the data. This also ensured that the models were not exposed to the 20% test data set aside at the beginning, to not bias the final models toward the test set. We selected the most important features that occurred in all 5 models. The gain was used as the measure of feature importance (i.e., the improvement in accuracy of classification attributable to a feature) (https://xgboost.readthedocs.io/en/stable/tutorials/model.html).[20]

Refitting XGBoost and selecting final models

We then used these top features to create new document–concept and concept models, again using the parameters from Appendix 1, Supplementary Table 1. Interpretability of our algorithms was a key study objective. We used a new technique[22] to compute SHapley Additive exPlanations (SHAP) values on trees, called TreeExplainer.[23] If a feature has a large positive SHAP value for a given patient, it indicates that the feature makes a positive finding of hypertension much more likely, with a large negative SHAP value indicating the opposite. We selected only those that were in the top 20 most important features across all folds, for both sets of models, to remove spurious features. We chose the top 20 as this appeared sufficient to capture the most relevant features, because feature importance decayed rapidly for both sets of models. Finally, we used these results to suggest a simpler concept search strategy for case identification and provide insights for future chart review.

Statistical analysis

We characterized the study cohort based on age, sex, type of admission (i.e., medical v. surgical) and hypertension status (from chart review). We determined the performance of the machine-learning–based algorithms across each fold of cross validation on the training, validation and testing sets separately. We calculated SHAP values for each of the features and used these to assess the importance of each feature to the final prediction. To determine the availability of each kind of EMR document, we computed the proportions of admissions that had each different document name available. We applied our hypertension phenotype to the extracted EMR text and the ICD definition to the extracted DAD records. We calculated sensitivity, specificity, PPV and NPV using the chart review hypertension status as a reference standard.

Ethics approval

This study was approved by the Conjoint Health Research Ethics Board at the University of Calgary (REB15-0790).

Results

Cohort characteristics are presented in Table 1. The median age of patients was 62 years, about half of the cohort were female (50.3%) and almost half had hypertension (48.5%). Most hospital admissions were nonsurgical (*n* = 1939, 63.8%) compared with surgical (*n* = 1101, 36.2%).

Machine learning model training

We split the patient data randomly: 80% (*n* = 2432) into a development (training) set, with the remaining 20% (*n* = 608) held out to test the performance of the final models, and

Table 1: Characteristics of the study cohort

Variable	No. (%) of patients*		
	All *n* = 3040	With hypertension† *n* = 1474	No hypertension† *n* = 1566
Age, yr; median (IQR)	62 (48–76)	71 (61–82)	52.5 (38–65)
Sex, female	1529 (50.3)	710 (48.2)	819 (52.3)
Admitted for surgery	1101 (36.2)	482 (32.7)	619 (39.5)

Note: IQR = interquartile range.
*Unless specified otherwise.
†We determined hypertension status by chart review.

Table 2: Performance of initial XGBoost document–concept models and concept models

Data used for validation	Training data*				Validation data*			
	Sensitivity, %		PPV, %		Sensitivity, %		PPV, %	
	DC	C	DC	C	DC	C	DC	C
Fold 0	90	100	93	100	89	89	88	89
Fold 1	85	100	94	100	86	92	94	93
Fold 2	90	100	94	100	87	91	90	91
Fold 3	90	100	94	100	88	91	92	92
Fold 4	89	100	94	100	91	94	91	89

Note: C = concept model, DC = document–concept model, PPV = positive predictive value.
*Folds 0 and 1: patients in the training (*n* = 1945) and validation (*n* = 487) cohorts; and folds 2, 3 and 4: patients in the training (*n* = 1946) and validation (*n* = 486) cohorts.

stratified by hypertension status to ensure there were about the same proportion of patients with hypertension in both sets. The performance of the initial document–concept and concept models are shown in Table 2, where each row is the fold used to calculate that model's performance, with the remaining 4 folds used to train that model. The concept models seemed to overfit on the training data but had similar performance to the document–concept models on the validation data. We found that all the models performed relatively well on the validation data, with sensitivities and PPVs close to 90% throughout. After limiting to the top 20 most important features in each fold, 10 document–concept and 8 concept features remained.

To evaluate how features affected the classification of each patient in the training set, we show the relation between feature and SHAP values for the concept model in Figure 2 and the document–concept model in Figure 3, where a larger SHAP value means a higher likelihood of classifying the patient as having hypertension. Figure 2 shows that the concept for hypertension (C0020538) is the most important feature in the concept model and is the only feature with strong negative classification results when absent. In Figure 3, all but 1 of the features in the document–concept model involve hypertension, which amounts to a ranked set of documents to search for hypertension documentation, with the best document to search being "Surgical Assessment and History — Nursing." The predominance of the hypertension concept in determining hypertension status showed that a concept search for C0020538 could also perform well and would have the benefit of being simpler to implement. These concepts are still extracted from cTAKES, and exclude those flagged by cTAKES as being negated (e.g., does not have hypertension) or refer to someone other than the patient (e.g., family history of hypertension).

Document availability

The availability of documents identified by the document–concept model are important in selecting documents that have a reasonable likelihood of being present. In order of importance of the identified feature: Surgical Assessment and History — Nursing was available in 36.7% of admissions; Nursing Transfer Report — Emergency Department to Inpatient was available in 57.9% of admissions; Discharge Summary — Medical was available in 41.6% of admissions; Nursing Transfer Report — Postanesthesia Care Unit to Inpatient was available in 38.1% of admissions; Adult Triage Note was available in 62.7% of admissions; Pharmacy Care Plan was available in 28.0% of admissions; Multidisciplinary Progress Report was available in 99.1% of admissions; History and Physical was available in 12.1% of admissions; and Discharge Summary was available in 15.5% of admissions.

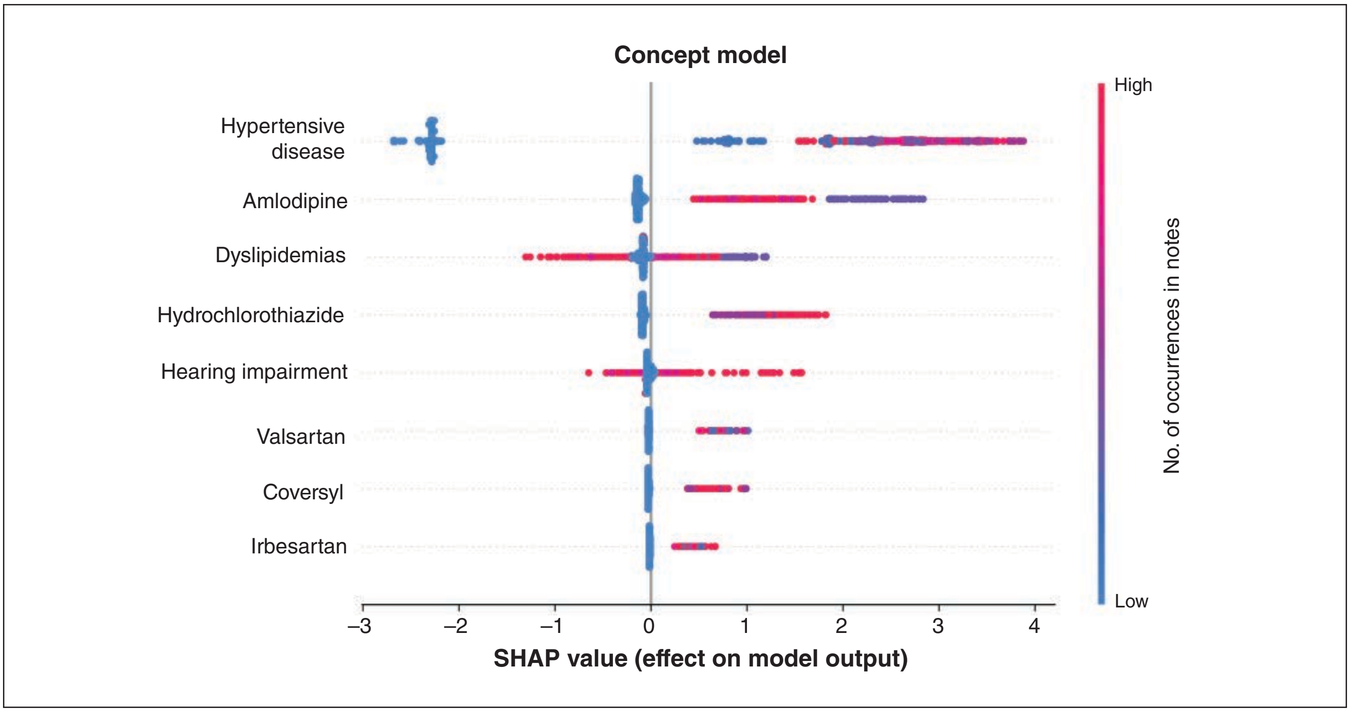

Figure 2: SHapley Additive exPlanations (SHAP) values for the final concept model. Values are for each patient in the training set (n = 2432), by model feature. Each dot represents a patient, with the SHAP value on the x-axis and feature value given by its colour. The higher the SHAP value the more likely the patient will be classified as having hypertension. The hypertension concept C0020538 is the most predictive feature, and the only one for which a low count results in a significant likelihood of the patient not being classified as having hypertension.

Model performance

Table 3 provides the results of the final machine-learning models, as well as the concept search algorithm and the ICD-10-CA algorithm, on the 20% of held out test data. The EMR algorithms had much higher sensitivities and NPVs than the ICD algorithm across all stratifications. This is offset by slightly worse PPVs, which are still above 90% for all groups except the youngest 2 age stratifications, where it drops as low as 82%. The youngest age stratification is the only place where the ICD algorithm has a lower PPV than the EMR algorithms. The ICD algorithm also has a higher specificity than the EMR algorithms, which are still above 90% for all groups except the oldest age stratification, where they drop as low as 87%. In general, we see that the concept search algorithm has quite comparable performance to the machine-learning algorithms.

Interpretation

We found hypertension could be accurately identified in an inpatient population using UMLS concepts extracted from EMR clinical notes, with a higher sensitivity than using a validated ICD algorithm. This shows that an EMR hypertension phenotype could be used in place of ICD case definition for health services research or be used to enhance existing administrative databases such as the DAD. Electronic medical record phenotypes like this also have the potential to be implemented directly in EMR systems to aid in clinical decision-making.

Our results were similar to those reported in a 2017 study involving the health records of 631 patients followed for hypertension status at Vanderbilt University School of Medicine,[24] where a hypertension case identification algorithm with a sensitivity of 96.6% and a PPV of 93.4% was developed. However, this algorithm also used primary care encounters, ICD codes and vital signs, as well as textual information.

Our work also provides insight into the underlying documentation of EMR data. The availability of documentation will vary greatly with the type of visit. As we noted earlier, only 54% of surgical visits contained a discharge summary compared with 86% of nonsurgical visits. Our document–concept algorithm showed that hypertension was documented most reliably in Surgical Assessment and History — Nursing followed by Nursing Transfer Report — Emergency Department to Inpatient, which occurred in 37% and 58% of admissions, respectively. Canadian coders are not required to review these nursing documents and only review physician documentation.[25] In hospitals, nurses check patient blood pressures and document them in nursing notes, and they also collect patients' daily clinical information. Thus, our EMR-based method could be automated, which could avoid potential bias associated with coding guidelines and practice.[26] This has the potential to improve ICD databases with minimal cost.

The presented EMR methods have various applications in clinical and research contexts (e.g., measuring and monitoring health system performance, cohort selection for research

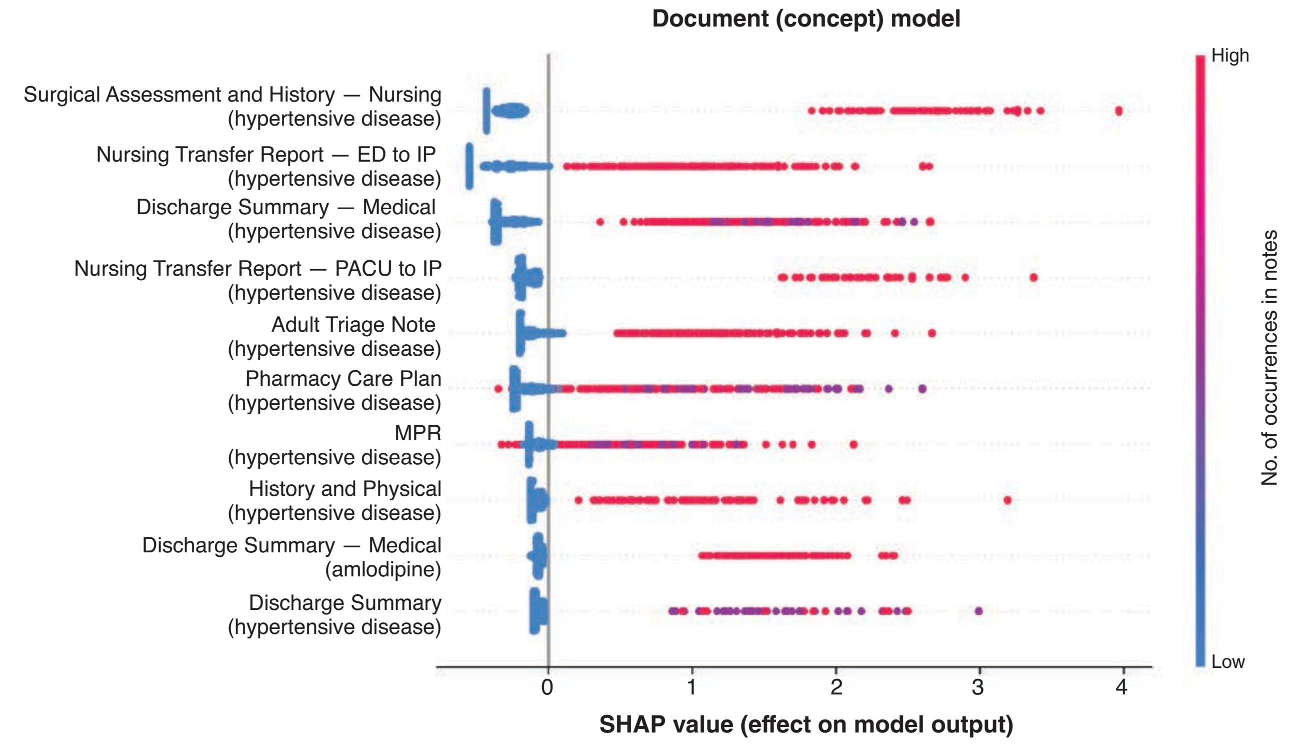

Figure 3: SHapley Additive exPlanations (SHAP) values for final document-concept model. Values are for each patient in the training set (*n* = 2432), by model feature. Each dot represents a patient, with the SHAP value on the *x*-axis, and feature value given by its colour. The higher the SHAP value, the more likely the patient will be classified as having hypertension. All the features represent different places to search for the hypertension concept C0020538, except the second to last, which looks for the amlodipine concept (C0051696) in the Discharge Summary — Medical document. Note: ED = emergency department, IP = inpatient, MPR = multidisciplinary progress report, PACU = postanesthesia care unit.

Table 3: Stratified validity scores across population characteristics for classification models document–concept model/concept model/concept search/validated *International Classification of Diseases* algorithm

	Sensitivity, %				Specificity, %				PPV, %				NPV, %			
Characteristic	DC	C	CS	ICD	DC	C	CS	ICD	DC	C	CS	ICD	DC	C	CS	ICD
Validation cohort (*n* = 608)	95*	91*	95*	47	92	93	92	98†	91	93	92	97†	95*	91	95*	66
By age, yr																
< 45 (*n* = 123)	100*	100*	100*	29	98	97	97	99†	88†	82	82	80	100*	100*	100*	92
45–64 (*n* = 206)	87	90*	90*	42	92	90	91	98†	87	84	85	94†	92	93	94†	74
> 64 (*n* = 279)	91	96	97†	50	88	87	87	97†	95	95	95	98†	79	89	90†	42
By admission type																
Surgical (*n* = 213)	90	91*	91*	44	94	93	93	99†	92	90	90	98†	92	93*	93*	70
Nonsurgical (*n* = 395)	91	96	97†	48	93	91	92	98†	93	92	92	96†	91	96*	96*	64
By sex																
Female (*n* = 302)	90	94*	94*	45	94	92	92	98†	92	91	91	95†	92	95*	95*	68
Male (*n* = 306)	91	95	96†	48	93	91	92	99†	93	92	93	97†	91	94	95†	64

Note: C = concept model, CS = concept search, DC = document–concept model, ICD = validated *International Classification of Diseases* algorithm, NPV = negative predictive value, PPV = positive predictive value.
*Highest performing models (tied).
†Highest performing model.

studies and surveillance). Although hypertension is most often diagnosed in a primary care setting, patients are admitted to hospital when the severity of the condition worsens. Therefore, identifying hypertension in an inpatient setting, without relying on primary care data, is essential for the assessment of health system performance. The machine-learning approach we present herein could also be applied to identifying other conditions in inpatient EMR data, which may not have as straightforward documentation. Our methods could be used with a common data model such as Observational Medical Outcomes Partnership.[27,28] This model would make use of a NOTE_NLP table where CUIs, their annotations and the document names are referenced.

Limitations

We used only inpatient documentation and are aware that hypertension is largely managed in outpatient settings. However, our study was aimed at developing EMR-based hypertension case identification to overcome undercoding issues in professionally coded ICD databases. Our reference standard identified cases based on clinician documentations and did not rediagnose hypertension based on charts. Although blood pressure measurements are part of diagnoses, the criteria can vary across countries, including cut-off values for blood pressure to define hypertension.[29,30] Therefore, clinical rule-based algorithms may not be as robust when performing case identification in other contexts. Finally, we have not conducted external validation of our algorithm using data from other jurisdictions. This type of external validation study between multiple systems may be feasible using common data models, such as Observational Medical Outcomes Partnership.[27,28]

Conclusion

We have leveraged EMR clinical notes to create a case identification algorithm for inpatients. We used machine-learning models to identify the most relevant concepts and documents to evaluate in EMRs and used those insights to create a simpler concept search case identification algorithm. This algorithm has the potential to improve the quality of hospital discharge abstract administrative data and also to provide a tool to measure rates of hospital admission for hypertension for system performance measurement and monitoring. The machine-learning models also provide insights into EMR documentation for future research, fulfilling the iterative feedback goal of a learning health system.

References

1. Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. *Sci Transl Med* 2010;2:57cm29.
2. Lee S, Doktorchik C, Martin EA, et al. Electronic medical record-based case phenotyping for the Charlson conditions: scoping review. *JMIR Med Inform* 2021;9:e23934.
3. Quan H, Li B, Duncan Saunders L, et al.; IMECCHI Investigators. Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. *Health Serv Res* 2008;43:1424-41.
4. Peng M, Chen G, Lix LM, et al.; Hypertension Outcomes Surveillance Team. Refining hypertension surveillance to account for potentially misclassified cases. *PLoS One* 2015;10:e0119186.
5. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. *Lancet* 2021;398:957-80.
6. Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care* 2005;43:1130-9.
7. Lucyk K, Tang K, Quan H. Barriers to data quality resulting from the process of coding health information to administrative data: a qualitative study. *BMC Health Serv Res* 2017;17:766.
8. Tang KL, Lucyk K, Quan H. Coder perspectives on physician-related barriers to producing high-quality administrative data: a qualitative study. *CMAJ Open* 2017;5:E617-22.
9. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res* 2004;32:D267-70.
10. Lee S, Xu Y, Apos Souza AGD, et al. Unlocking the potential of electronic health records for health research. *Int J Popul Data Sci* 2020;5:1123.
11. Lee S, Li B, Martin EA, et al. CREATE: a new data resource to support cardiac precision health. *CJC Open* 2020;3:639-45.
12. Xu Y, Lee S, Martin E, et al. Enhancing ICD-code-based case definition for heart failure using electronic medical record data. *J Card Fail* 2020;26:610-7.
13. Eastwood CA, Southern DA, Khair S, et al. Field testing a new ICD coding system: methods and early experiences with ICD-11 Beta Version 2018. *BMC Res Notes* 2022;15:343.
14. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015;162:55-63.
15. Lachenbruch PA. On the sample size for studies based upon McNemar's test. *Stat Med* 1992;11:1521-5.
16. Quan H, Smith M, Bartlett-Esquilant G, et al. Mining administrative health databases to advance medical science: geographical considerations and untapped potential in Canada. *Can J Cardiol* 2012;28:152-4.
17. Perrier A. *Effective amazon machine learning: machine learning in the cloud*. Birmingham (UK): Packt Publishing; 2017.
18. Dangeti P. *Statistics for machine learning: techniques for exploring supervised, unsupervised, and reinforcement learning models with Python and R*. Birmingham (UK): Packt Publishing; 2017:1-442.
19. Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc* 2010;17:507-13.
20. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: Association for Computing Machinery; 2016:785-94. Available: https://dl.acm.org/doi/10.1145/2939672.2939785 (accessed 2022 July 21).
21. Mohri M, Rostamizadeh A, Talwalkar A. *Foundations of machine learning, second edition*. Cambridge (MA): MIT Press; 2018:1-504.
22. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell* 2020;2:56-67.
23. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Guyon I, Von Luxburg U, Bengio S, et al. *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017)*; 2017 Dec. 4–9; Long Beach (CA), Red Hook (NY): Curran Associates Inc.:4768-77. Available: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html (accessed 2022 Apr. 30).
24. Teixeira PL, Wei W-Q, Cronin RM, et al. Evaluating electronic health record data sources and algorithmic approaches to identify hypertensive individuals. *J Am Med Inform Assoc* 2017;24:162-71.
25. *Canadian Coding Standards for Version 2018 ICD-10-CA and CCI*. Ottawa: Canadian Institute for Health and Information; 2018:1-767.
26. Saposnik G, Redelmeier D, Ruff CC, et al. Cognitive biases associated with medical decisions: a systematic review. *BMC Med Inform Decis Mak* 2016;16:138.
27. Hripcsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform* 2015;216:574-8.
28. Overhage JM, Ryan PB, Reich CG, et al. Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc* 2012;19:54-60.
29. Garies S, Hao S, McBrien K, et al.; Hypertension Canada's Research and Evaluation Committee. Prevalence of hypertension, treatment, and blood pressure targets in Canada associated with the 2017 American College of Cardiology and American Heart Association blood pressure guidelines. *JAMA Netw Open* 2019;2:e190406.
30. Khera R, Lu Y, Lu J, et al. Impact of 2017 ACC/AHA guidelines on prevalence of hypertension and eligibility for antihypertensive treatment in United States and China: nationally representative cross sectional study. *BMJ* 2018;362:k2357.

Affiliations: Centre for Health Informatics (Martin, D'Souza, Lee, Eastwood, Quan) and Department of Community Health Sciences (Eastwood, Quan), Cumming School of Medicine, University of Calgary; Alberta Health Services (Martin, D'Souza, Lee), Calgary, Alta.; Department of Medicine, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alta.

Contributors: Elliot Martin performed most of the analysis and writing. Adam D'Souza assisted with the analysis and writing. Seungwon Lee and Chelsea Doktorchik provided subject matter expertise and assisted with writing. Cathy Eastwood and Hude Quan were responsible for study design and assisted in the writing. All of the authors reviewed the manuscript critically for important intellectual content, gave final approval of the version to be published and agreed to be accountable for all aspects of the work.

Funding: This work was supported by a Canadian Institutes of Health Research Operating Project Grant (201809PJT-409935-HS1-CBBA-114817).

Data sharing: The data sets analyzed in this study are not publicly available because of the potential for identifying information to be exposed in the clinical notes. Working with the data is possible through collaboration with the Centre for Health Informatics and Alberta Health Services. The Python code used to perform this analysis can be found at the GitHub repository (https://github.com/centre-for-health-informatics/Hypertension-Case-Identification, doi: 10.5281/zenodo.4543942) under an open Massachusetts Institute of Technology license.

Acknowledgements: The authors thank Yuan Xu and Sang Min Lee for their helpful comments on the original manuscript, as well as Robin Walker for her help in revising it.

Supplemental information: For reviewer comments and the original submission of this manuscript, please see www.cmajopen.ca/content/11/1/E131/suppl/DC1.