ACAM, 2023, volume 10, issue 1

Annals of Clinical and Analytical Medicine

Applications of Natural Language Processing in Cardiology using Text Clinical Data: A Systematic Review

Hamdi. A. Alhakimi* (1) , Tayseer Elsiddig Magzoub (2)

(1) Epidemiologist, Community Medicine Specialist, Utrecht University, The Netherlands.
(2) Computer Engineering Master, Arab Open University, Riyadh, Saudi Arabia.

Received 15/6/2023; revised 8/9/2023; accepted 15/9/2023

****Corresponding author:*** Dr.hamdi.hakimi@gmail.com

Abstract

A low survival rate of heart failure (HF) is attributed to the under-diagnosis due to lack of the diagnostic reference standard. Heart failure usually is not well-documented in the administrative databases due to inconsistency in use of diagnostic codes and inter-examiner variability. The majority of EHR databases can export data for certain patients' characteristics, such as demographics and lab results, in a structured and analysis-friendly format. However, a lot of clinical data are stored in text and unstructured format. The use of unstructured clinical text data can substantially enhance both discission-making and clinical research. A manual extraction of unstructured data is time- and money- consuming process, hence using NLP algorithms with automatic extraction and classification could enhance efficiency and accuracy of the process.

This review aimed to highlight the literature that addressing application of NLP in the analysis of clinical text data related to diagnosis and prognosis of cardiovascular diseases. A multiple-term search strategy was used in PubMed and resulted in 53 studies, while only 20 studies used NLP techniques to handle text data related to HF. The included studies used NLP in different clinical purposes such as clinical features extraction, HF classification, and prediction of various HF outcomes. Early detection of HF symptoms was achieved in many studies and sometimes a median time of 6 months was found between a symptom reporting and the clinical diagnosis. Not only symptoms were extracted, characteristics of self-management, social determinants, and home-care were successfully identified by NLP techniques. Ejection fraction in clinical notes was used mainly to determine the type and severity of HF and it was associated with very good performance of NLP classifiers. Using of semi-structured clinical data, such as radiological reports, were usually associated with a better performance than using unstructured data, such as nurse notes. However, a combination of different types of data, particularly those supported by expert-knowledge, showed a promising results in HF diagnosis or prognosis. Using NLP techniques in the future can reduce underestimation of HF, particularly, when computer-extracted features and expert-optimized concepts are combined.

Keywords: Heart Failure, Natural Language Processing, Diagnosis, Prognosis, Text data.

Introduction

The growing phenomenon of aging, world widely, poses more burden on healthcare systems with an estimation of 56% increase in the number of over-60 years old people by 2030 (1). Cardiovascular diseases are age-related conditions that ranked as the most common leading cause of death globally (2). Heart failure (HF) is one of the most important cardiovascular diseases with estimated 64.34 million cases over the world (3), of them 51% are considered to be severe (4).

Heart failure is a growing health problem with lifetime risk of 20% estimated at age 40 years old with five-year survival between 20% to 50% (5-7). In addition to poor quality of life, patients with HF have a high readmission rate which could be as high as 20-30% for 3-month period (8). The low survival rate is attributed to the under-diagnosis of HF due to lack of the diagnostic reference standard. Heart failure usually is not well-documented in the administrative databases due to inconsistency in use of diagnostic codes and inter-examiner variability. In a systematic review aimed to evaluate the code validity of International Classification of Diseases (ICD) of HF in Electronic Health Record (EHR) databases, only 37% of the included studies used standardized criteria such as Framingham criteria, which were limited by the quality of reporting (9). The rest of the included studies depends on reviewing patients' chart or patient-reported outcomes, which have many data-quality issues. The manual review of EHR databases for identification of complex disease, such as HF, is a time-consuming process and usually required trained staff with proper data-quality management. For instance, Framingham criteria for diagnosis of HF include many major and minor phenotypes. The major criteria include 4.5 kg weight variation within 5 days of hospitalization period, jugular venous distension, orthopnea, radiographical cardiomegaly dyspnea on rest, hepatojugular reflux, radiographic pulmonary oedema, S3 gallop, paroxysmal nocturnal dyspnea, and pulmonary basilar rales. Moreover, there are minor criteria of HF such as night cough, dyspnea in exertion, hepatomegaly, lower limbs oedema, and the bilateral pleural effusion. Hence, many investigators attempted to automate the EHR review using different methods including rule-based programing methods, natural language processing algorithms (NLP), or a combination of both.

The majority of EHR databases can export data for certain patients' characteristics, such as demographics and lab results, in a structured and analysis-friendly format. However, a lot of clinical data are stored in text and unstructured format. In the era of big data and a rapid development of machine learning, innovations may efficiently and comprehensively use text clinical data, which are available in electronic health records (EHR). Narrative documentation of clinical events is considered as the most natural and expressive manner to report clinical history, clinical presentations, and discharge notes. Due to their unstructured nature, text data are hard to be analyzed in order to draw beneficial conclusions aiding in the diagnosis, prognosis, and treatment. However, NLP algorithms have been found to be powerful tools in the analysis of text data in both quantitative and qualitative approaches (10, 11). The NLP is a branch of machine learning that deal with text data to achieve different classification and prediction tasks and try to mimic human-like language processing (12).

Some studies used NLP for extraction of Framingham criteria from unstructured clinical notes. Processing of unstructured text data usually starts by tokenization, followed by stemming, lemmatization, removal of stop words, and vectorization (Figure 1). The tokenization is the process of transforming the text into tokens where each sentence, word, comma or stop word could be a token. Stemming is finding the root of the word which is equal or smaller form of the word, while lemmatization is referred to removal of unnecessary part of the words such as removing "s" from "episodes". Word vectorization or word embedding is the process of transforming the text into a numerical vector of data. In python, it could be obtained by transforming the text into "dictionary" form of data. Word vectorization includes many actual

methods, such as N-gram and Bag-of-Words. The N-gram vectorization considered the sequence of n-words such as 2-gram or 3-gram which consists of 2 or 3 words, respectively. Bag-of-Words count the word occurrence in each document without preserving the order of the words. Another common vectorization method is a term frequency-inverse document frequency vectorization (TFIDF), which penalizes common words in the document such as "for", "of", and "in". Thus TFIDF could be considered as a method to give more weight to the rare words which are usually more important than common words. Some data processing techniques are more common in the clinical data such as negation. The concept recognition means an identification of the "feature presence" in the text, such as "ankle oedema" or "pleural effusion", while the concept negation means identification of negation of the feature such as " no oedema" or "pleural effusion is ruled out" .

Use of unstructured text data from EHR, that previously were considered as non- or minimally useful data, can substantially enhance both discission-making and clinical research. This is particularly true for cardiovascular diseases due to their longitudinal nature with enormous and continuous stream of text data. Manual transformation of text data into structured data is time- and money- consuming process, hence using NLP algorithms with automatic extraction and transformation will enhance efficiency and accuracy of the process. This review aims to highlight the literature that addressing application of natural language processing in the analysis of clinical text data related to diagnosis and prognosis of cardiovascular diseases.

Methods

A search strategy was used in PubMed to identify studies used NLP to handle text data related to HF. This search strategy was as following: (heart failure OR myocardial failure OR HFrEF OR HFmrEF OR HFpEF OR ejection fraction OR EF) AND (NLP OR natural language processing) AND (text data OR medical records OR electronic health records OR electronic patient data OR electronic patient records) AND (mortality OR death OR incidence OR risk OR survival OR quality of life OR recovery OR readmission OR diagnosis OR length of stay OR relapse OR heart arrest OR prognosis OR pulmonary oedema).

Results

The electronic search resulted in 53 studies which were subjected to a primary screening by reading their abstracts. Based on the abstract, only 20 studies used NLP techniques to handle text data related to HF. The included studies used NLP in different clinical purposes such as clinical features extraction, HF classification, and prediction of various HF outcomes. Moreover, some studies used NLP to investigate various public health problems such as assessment of communication failure, evaluation of self-management, and identification of important social determinants of HF (see flow chart in Figure 2).

Regarding methods of NLP that were used in the included studies, many studies used ready-made NLP tools, while other studies developed their own NLP algorithms using algorithms such as neural networks (13) (14), Naïve Bayesian (15), support vector machine (16), forest plot (16) , and gradient boosting (17). As the majority of the included studies were conducted by clinicians, they are mainly focused on clinical importance and only few studies reported adequate technical details about NLP algorithms (16) (18) (13) (19). Generally, information about pre-processing of data, algorithm infrastructure, tuning of model hyperparameters, details of model training and validation, were poorly documented in the studies. The input data were either structured data such as patients' demographic, unstructured data such as clinical notes, or a combination of both. Furthermore, performance metrics of NLP models were not consistent across all studies, as some studies used recall and precision, others reported accuracy and F1, and some prognostic studies calculated area under the curve (AUC) for receiver operating curve. However, most studies reported precision as a performance measure, also known as positive predictive value, which enable comparisons between the algorithm performance across the included studies.

Discussion

In the literature, we found three main uses of NLP in dealing with HF clinical data including the extraction of important clinical features or criteria of HF, the classification of patients into different diseases categories, and the prediction of HF-related mortality or readmission rates.

Extraction of HF features by NLP:

In the clinical data, with well-organized medical vocabularies, many algorithms were developed specifically to process the clinical data, such as Spark NLP-healthcare (20). As other ML algorithms, extraction of features from text data to be used for algorithm training is mainly based on the clinical knowledge about the disease (21). In regards to HF, Framingham criteria were the most commonly extracted features of HF among the included studies.
In the majority of the include studies, the NLP extracted criteria were compared to manually extracted criteria which were identified by trained reviewer as a gold standard. However, in some studies, the main use of NLP techniques was only to extract features, which then were fed to other classification techniques such as statistical modelling or rule-based programming (22).

For instance, Moore and colleagues developed an NLP tool using python and open-source software called cTAKES (22). They used an open-source NLP tool to extract the major and minor HF criteria from clinical notes related to 394 hospitalizations including emergency notes, admission reports, radiological reports, and discharge summaries. They found that the prevalence of the Framingham criteria in clinical notes was 52% and 55.8% using NLP and manual review, respectively (22). These findings show the ability of NLP techniques to achieve a comparable performance to that of human expert. The NLP techniques can differentiate between cases and non-cases in early stage of the disease as demonstrated by a retrospective extraction of symptoms at the primary healthcare setting. Hence, among 50,000 patients, a significant difference in the prevalence rate between cases and non-cases was detected with 62% versus 24% (18).

By focusing only on patients with complains, a hypothesis-free NLP approach was used to identify manifestations of HF (23). Manifestations of decompensated HF such as dyspnea, acrocyanosis, and respiratory failure were common with a prevalence of 98.1%, 43.2%, and 41.9% of the patients' records (23). Another interesting finding of this retrospective analysis is early detection of the symptoms with a median time of 6 months between the symptom reporting and the clinical diagnosis. Not only symptoms of HF were extracted from clinical databases, an extraction of features related to poor self-management of HF were also identified with 0.86 precision rate using NLP algorithms called Nimble Miner (24). In 2.3 millions of clinical notes for 67,683 patients, the highest precision was found for a symptom called "confusion", while the lowest was for a symptom "nausea". For poor self-management, the best precision was for terms "unspecified non-adherence" or "did not check blood pressure", while the lowest precision was for "missed follow-up" (24). These terms could differ from hospital to another or form country to country, which highlighted the importance of specific text data analysis for each situation.

Moreover, NLP techniques could identify the importance of each extracted terms with regards to the disease of interest. Wang and colleagues extracted 32 terms from HER database including two terms from demographics, two from vital signs, four from comorbidities, and one from clinical history (25). The investigators identified top four important terms that directly associated with CHF, including "heart failure", "congestive heart", "congestive heart failure" and "chf". Interestingly, age was highly related to hospitalization rate of CHF. Other discriminant terms for CHF were "glucose" and "blood pressure" which are clinical terms related to the presence of diabetes mellitus and hypertension (25). Hence, the evaluation of feature importance could enhance our clinical knowledge about the disease by utilization of the large amount of neglected text data. Some studies focused on certain types of clinical data such as radiological reports which could be subjected to different steps of

text analysis. A convolutional neural network was used only for extraction of 14 common findings in thoracoabdominal CT such as aortic aneurysm, ascites, and atelectasis. The accuracy of the extraction process was good with precision range between 0.86 and 0.97, in reference to manual feature identification. Again, radiological reports seem to have a high precision in the extraction of the clinical features related to HF (13). Among 39,000 chest x-ray reports, an NLP data extractor was developed to identify concepts related to CHF (26). The investigators called the NLP tool as REX which started by sectioning the reports before application of usual tokenization and lemmatization of text data. The authors found very optimistic results, as the precision was perfect for "Kerley B lines" (100%) and very high for "cardiomegaly", "prominent pulmonary vasculature", and "pleural effusion" phrases (99%), while for CHF phrases, the precision was 95%. A very high precision may indicate a problem of overfitting but focusing on one type of semi-structure data, such as radiological reports, could markedly improve the performance of NLP algorithms. These findings may suggest that NLP models should be specific for type of data and should be trained on each type of clinical notes (13). It is logical that clinical notes with semi-structured format like radiological reports facilitate identification and extraction of the required clinical features.

Heart Failure as an NLP classification problem:

Almost all included studies evaluated the performance of NLP algorithms in reference to manually extracted data, as a gold-standard. The majority of the studies identify the presence of features related to HF from text data, while few studies attempted to classify HF into different diagnostic categories such as HF with preserved EF (HFpEF) or HF with reduced ejection fraction (HFrEF). Moreover, some studies used a combination of HF-concepts and ICD-related terms to train and validate NLP classifiers. An example of features-based classification is a study conducted by Moore and colleagues who compared HF criteria, extracted by cTAKES, to manually extracted data and reported the precision of each criterion (22). After validation of the performance in clinical note of 406 new hospitalizations, the overall precision was 84.4% for all 14 HF criteria. They found that cardiomegaly and dyspnoea had the best precision with 96.7% and 94.5%, while S3 gallop and hepatojugular reflux had the lowest precision with 11.8% and 0.01%, respectively (22). It seems like the performance of the classifiers is dependent on the extracted features. However, another study found that the type of NLP techniques are more crucial in the classification of HF. Linear NLP classifiers was used to identify patients with CHF, based on features extracted from clinical notes in Mayo Clinic. The authors validated NLP algorithms using 10-folds cross validation in reference to a physician-annotated dataset. The algorithms identified key-terms such as ''cardiomyopathy'', ''heart failure'', ''congestive heart failure'', and ''fluid overload''. The performance of Naive Bayes classifier was found better than that in Perceptron neural network with (100 vs. 85%) recall (15).

As ICD is more systematic than HF-concept phrases, a search algorithm could easily identifies the cases using their specific keywords. In a new approach, three NLP methods were used to identify patients with HF (17). These methods are a keyword search algorithm which return ICD-related terms, HF-concept algorithm which is a gradient boosting (XGBoost) algorithm that was used to extract 6 important HF-features in order to use them in building interpretable classifiers, and a combination of both ICD and HF-concept algorithms. The findings revealed that ICD algorithm had the highest precision (92.4%) with low recall (57.4%), while HF-concept algorithm had a slightly lower precision (88.9%) but with high recall (80%). Combining both algorithms were not likely to improve the performance as it increased the recall (83.3%) but decreased the precision (83.3%) (17). However, combining ICD search with sophisticated NLP techniques, such as Bidirectional Gated Recurrent Unit Neural Network (BGRU), resulted in good performance particularly for four-character ICD-10 codes which ranges from 0.87 to 0.98 (14). Moreover, using medications' terms to detect cardiovascular diseases was associated with high performance of BGRU, but with a higher risk of over-classification due to prescription of similar drugs for different cardiovascular diseases. For instance, amlodipine and perindopril were prescribed for

treatment of hypertension but they are prescribed sometimes for prevention of heart failure (27). In regards to features related to the ejection fraction, the type or severity of HF was the focus of the majority of the studies. The ejection fraction is an important indicator of heart failure that have been used by many studies for diagnosis of risk stratification of HF. Kim and colleagues used 18,397 records, including different types of clinical and radiological reports, to identify LVEF. They used sophisticated approach with three extraction modules and trained the algorithms to generate labels for 4 concepts including LVEF, LVSF, quantitative measurement, and qualitative values. They observed a higher performance in semi-structured reports than that in unstructured reports and they achieved a recall of 0.98 and a precision of 0.99 (28). Another study compared measurements of EF between radiological reports and electrocardiogram records in order to identify the type of HF(29). The authors found six common terms in the reports of 89% of 706 patients with HF. These terms included "multi-organ", "CHF", "cardiac failure", "heart failure"," ventricular failure", and "LVF". When they combined these terms with ICD-9 codes, the performance was excellent with 99% recall rate. Identification of cases with HFpEF from a large dataset with different types of clinical notes was the aim of a study that conducted by Patel and colleagues (30). The precision of the identification of HFpEF patients was 96%, while recall was 88%. However, the performance was lower with precision and recall values of 75% and 86%, when the diagnosis was categorized as definite, probable, and likely cases with HFpEF (30).

A relation between certain predictors, such as liver fibrosis, and certain types of HF such as HFrEF and HFpEF, was investigated by So-Armah and colleagues (31). Ejection fraction was extracted from echocardiogram file using NLP techniques and it was categorized into >50%, 40-50%, and <40% which labelled as preserved, mid-range, and reduced ejection fraction. Among all predictors, only the association between liver fibrosis and HFpEF was significant with hazard ratio of 1.7 (95% confidence interval 1.3 to 2.3). In order to estimate the true number of patients with HF at a tertiary hospital, a comparison was made between ICD-based HF diagnosis and NLP-based definition of HF cases (19). The authors compared 3 approaches for search query including ICD-based queries, an initially-expert specified queries, and computer- and expert-optimized queries. The findings shows that using ICD criteria alone was responsible for underestimation of HF by 44% with range of 55% in a single year to 31% in all years. The detection rate changed per year which indicates a secular variations in HF coding practices. Using NLP techniques in the future could reduce underestimation of HF, particularly, when computer-extracted features and expert-optimized concepts are combined.

Prognosis of HF outcomes using NLP:

Natural language processing was used in the literature to predict HF outcomes using various input data in different setting. For instance, researchers investigated the effect of the poor nurse-physician communication on 30-day hospital readmission during home-healthcare among discharged patients with CHF (32). An NLP algorithm screened the text data of nurse notes in thousands of communication episodes and identified the presence of communication failure. They found that communication failure was associated with 32.6% increase in the mean readmission rate among high-risk patients (32). Hence, the NLP techniques could help in improving managerial issues in clinical practice. Health system research could get benefits from NLP by focusing in communication, satisfaction, and confidence issues of both patients and care providers. Moreover, Another study aimed to develop prognostic models for cardiovascular diseases based on social determinants of health (33).

The authors used an NLP tool called Moonstone which is developed to identify poorly documented variables in clinical data. On the validation phase, the overall precision of social determinants identification was 83%. Furthermore, a high level of precision (>90%) was obtained in identification of certain factors such as medication compliance, living alone, and depression (33). For psychological conditions like depression and quality of life, using NLP seems to be a very good idea. Semantic analysis of text data can identify the emotional characteristics of the text data which is so helpful in assessment of psychological

conditions (34). For early diagnosis and prognosis of HF among hospitalized patients, an automated real-time risk assessment NLP algorithm was developed (35). The NLP algorithm identified patients with heart failure from clinical reports and the outputs were added to the regression prediction model. For risk assessment, a very high precision (97.5%) was obtained with 8% added prediction value due to NLP outputs. The use of NLP as a real-time risk estimator could improve the predictive ability to obtain a better prognosis than that depends on classical statistical modelling. Other combinations of statistical techniques and machine learning methods were used to predict the risk of 30-day readmission due to HF (36). Three regression models were developed based on parameters from structured data, unstructured data, and a combination of all parameters. Structured data with potential predictors of HF were analyzed using regression analysis after imputation of missing data, while parameters were extracted from unstructured data using NLP methods. The performance of structured and combined models were similar with AUCs (0.65), while it was lower (0.52) for unstructured model (36). In the future, researcher should focus in using combination of structured and non-structured data for the same patients in order to gain the benefits of both types of data.

Not all NLP tools are complicated, Orange-3 is a user-friendly software that did not require technical coding skills. Kang and colleagues use Orange-3 to compare the utility of different types of clinical notes to forecast 30-day readmission rate among patients with HF (16).. The data were fed into 6 different ML algorithms including logistic regression, random forest, support vectors machine, Naïve Bayes, neural networks and k-nearest neighbor clustering. Interestingly, the model used nurses' notes had better performance than semi-structured templates of patients' discharges. Based on the best model which is a model of word-bagging with neural network, the area under the curve was 0.95 for the nurses' notes versus 0.74 for the discharge summaries. Hence, use of of semi-structured data is not always superior to unstructured data in NLP. Combinations of both structured and unstructured input data and different NLP methods seems to improve the ability to classify or predict health problems. In regards to limitations, some studies had signs of overfitting as the performance was close to the unity (100%), other studies showed over-classification when using general terms such as drugs prescription.

Conclusions

The NLP techniques showed ability to achieve a comparable performance to that of human expert, particularly in extraction of features related to HF. Early detection of HF symptoms was achieved in many studies and sometimes a median time of 6 months was found between a symptom reporting and the clinical diagnosis. Using NLP techniques could reduce underestimation of HF, particularly, when computer-extracted features and expert-optimized concepts are combined. Not only symptoms were extracted, characteristics of self-management, social determinants, and home-care were successfully identified by NLP techniques. Furthermore, NLP could identify the importance of extracted features or characteristics with regards to HF diagnosis or prognosis. Type of extracted features, such as cardiomegaly and pleural effusion, were found to be important in the classification of HF from text clinical data. Using of semi-structured clinical data, such as radiological reports, were usually associated with a better performance than using unstructured data, such as nurse notes. However, a combination of different types of data, particularly those supported by expert-knowledge, showed a promising results in HF diagnosis or prognosis.

Conflict of interests

The authors declared no conflict of interests.

References

1. Nations, U., Department of Economic and Social Affairs; Population Division. World Population Ageing 2015. 2013, Author New York, NY.
2. Mensah, G.A., G.A. Roth, and V. Fuster, The global burden of cardiovascular diseases and risk

factors: 2020 and beyond. 2019, American College of Cardiology Foundation Washington, DC. p. 2529-2532.

3. George, J., et al., How does cardiovascular disease first present in women and men? Incidence of 12 cardiovascular diseases in a contemporary cohort of 1 937 360 people. Circulation, 2015. 132(14): p. 1320-1328.

4. Lippi, G. and F. Sanchis-Gomar, Global epidemiology and future trends of heart failure. AME Med J, 2020. 5(15): p. 1-6.

5. Benjamin, E.J., et al., American heart association council on epidemiology and prevention statistics committee and stroke statistics subcommittee. Heart disease and stroke statistics-2018 update: a report from the American Heart Association. Circulation, 2018. 137(12): p. e67-e492.

6. Bleumink, G.S., et al., Quantifying the heart failure epidemic: prevalence, incidence rate, lifetime risk and prognosis of heart failure: the Rotterdam Study. European heart journal, 2004. 25(18): p. 1614-1619.

7. Koudstaal, S., et al., Prognostic burden of heart failure recorded in primary care, acute hospital admissions, or both: a population-based linked electronic health record cohort study in 2.1 million people. European journal of heart failure, 2017. 19(9): p. 1119-1127.

8. Young, J.B., The global epidemiology of heart failure. Medical Clinics, 2004. 88(5): p. 1135-1143.

9. Yancy Clyde, W., et al., ACC/AHA/HFSA focused update of the 2013 ACCF/AHA guideline for the management of heart failure. J. Am. Coll. Cardiol, 2017. 70: p. 776-803.

10. Tanguy, L., et al., Natural language processing for aviation safety reports: From classification to interactive analysis. Computers in Industry, 2016. 78: p. 80-95.

11. Crowston, K., E.E. Allen, and R. Heckman, Using natural language processing technology for qualitative data analysis. International Journal of Social Research Methodology, 2012. 15(6): p. 523-543.

12. Nadkarni, P.M., L. Ohno-Machado, and W.W. Chapman, Natural language processing: an introduction. Journal of the American Medical Informatics Association, 2011. 18(5): p. 544-551.

13. Pandey, M., et al., Extraction of radiographic findings from unstructured thoracoabdominal computed tomography reports using convolutional neural network based natural language processing. PloS one, 2020. 15(7): p. e0236827.

14. Sammani, A., et al., Automatic multilabel detection of ICD10 codes in Dutch cardiology discharge letters using neural networks. NPJ digital medicine, 2021. 4(1): p. 1-10.

15. Pakhomov, S.V., J. Buntrock, and C.G. Chute, Prospective recruitment of patients with congestive heart failure using an ad-hoc binary classifier. Journal of biomedical informatics, 2005. 38(2): p. 145-153.

16. Kang, Y., et al., The Utility of Nursing Notes Among Medicare Patients With Heart Failure to Predict 30-Day Rehospitalization: A Pilot Study. The Journal of Cardiovascular Nursing, 2021.

17. Xu, Y., et al., Enhancing ICD-code-based case definition for heart failure using electronic medical record data. Journal of Cardiac Failure, 2020. 26(7): p. 610-617.

18. Vijayakrishnan, R., et al., Prevalence of heart failure signs and symptoms in a large primary care population identified through the use of text and data mining of the electronic health record. Journal of cardiac failure, 2014. 20(7): p. 459-464.

19. Kaspar, M., et al., Underestimated prevalence of heart failure in hospital inpatients: a comparison of ICD codes and discharge letter information. Clinical Research in Cardiology, 2018. 107(9): p. 778-787.

20. Kocaman, V. and D. Talby, Spark NLP: natural language understanding at scale. Software Impacts, 2021. 8: p. 100058.

21. Spasic, I. and G. Nenadic, Clinical text data in machine learning: systematic review. JMIR medical informatics, 2020. 8(3): p. e17984.

22. Moore, C.R., et al., Ascertaining Framingham heart failure phenotype from inpatient electronic health record data using natural language processing: a multicentre Atherosclerosis Risk in Communities (ARIC) validation study. BMJ open, 2021. 11(6): p. e047356.

23. Nagamine, T., et al., Multiscale classification of heart failure phenotypes by unsupervised clustering of unstructured electronic medical record data. Scientific reports, 2020. 10(1): p. 1-13.

24. Chae, S., et al., Identifying heart failure symptoms and poor self-management in home healthcare: a natural language processing study. Stud Health Technol Inform, 2021. 284: p. 15-19.

25. Wang, Y., et al., NLP based congestive heart failure case finding: A prospective analysis on statewide electronic medical records. International journal of medical informatics, 2015. 84(12): p. 1039-1047.

26. Friedlin, J. and C.J. McDonald. A natural language processing system to extract and code concepts relating to congestive heart failure from chest radiology reports. in AMIA annual symposium proceedings. 2006. American Medical Informatics Association.

27. Rosendorff, C., et al., Treatment of hypertension in the prevention and management of ischemic heart disease: a scientific statement from the American Heart Association Council for High Blood Pressure Research and the Councils on Clinical Cardiology and Epidemiology and Prevention. Circulation, 2007. 115(21): p. 2761-2788.

28. Kim, Y., et al., Extraction of left ventricular ejection fraction information from various types of clinical reports. Journal of biomedical informatics, 2017. 67: p. 42-48.

29. Bielinski, S.J., et al., A robust e-epidemiology tool in phenotyping heart failure with differentiation for preserved and reduced ejection fraction: the electronic medical records and genomics (eMERGE) network. Journal of cardiovascular translational research, 2015. 8(8): p. 475-483.

30. Patel, Y.R., et al., Development and validation of a heart failure with preserved ejection fraction cohort using electronic medical records. BMC Cardiovascular Disorders, 2018. 18(1): p. 1-8.

31. So-Armah, K.A., et al., FIB-4 stage of liver fibrosis is associated with incident heart failure with preserved, but not reduced, ejection fraction among people with and without HIV or hepatitis C. Progress in cardiovascular diseases, 2020. 63(2): p. 184-191.

32. Pesko, M.F., et al., Home health care: nurse–physician communication, patient severity, and hospital readmission. Health services research, 2018. 53(2): p. 1008-1024.

33. Reeves, R.M., et al., Adaptation of an NLP system to a new healthcare environment to identify social determinants of health. Journal of Biomedical Informatics, 2021. 120: p. 103851.

34. Le Glaz, A., et al., Machine learning and natural language processing in mental health: systematic review. Journal of Medical Internet Research, 2021. 23(5): p. e15708.

35. Evans, R.S., et al., Automated identification and predictive tools to help identify high-risk heart failure patients: pilot evaluation. Journal of the American Medical Informatics Association, 2016. 23(5): p. 872-878.

36. Mahajan, S.M. and R. Ghani. Combining Structured and Unstructured Data for Predicting Risk of Readmission for Heart Failure Patients. in MedInfo. 2019.

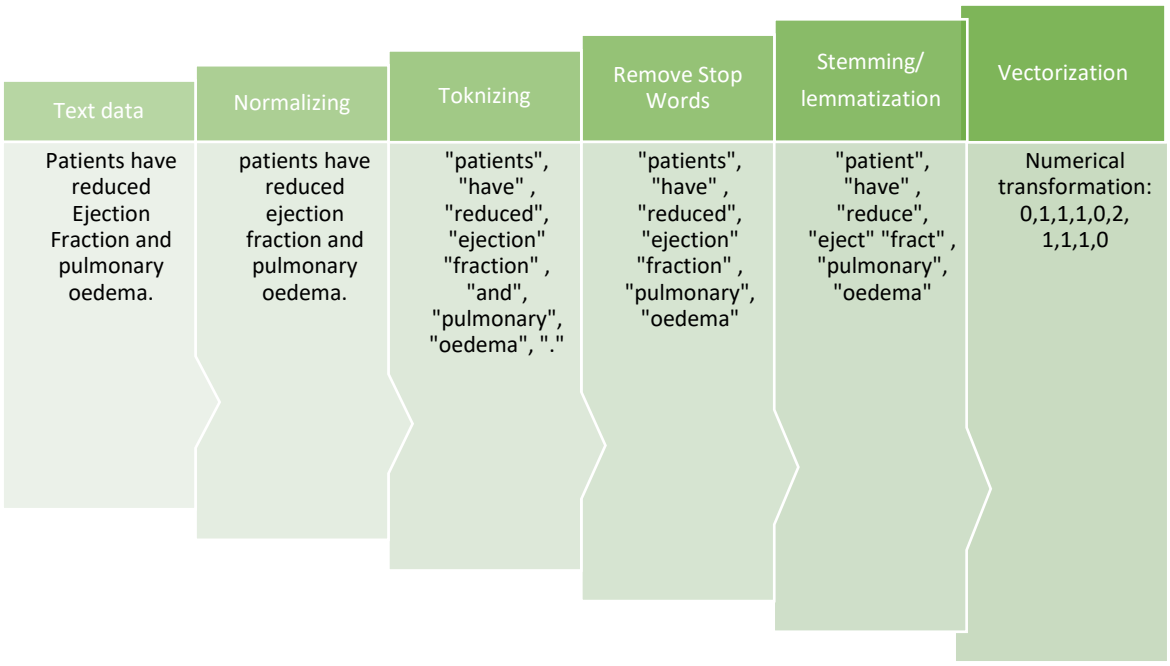
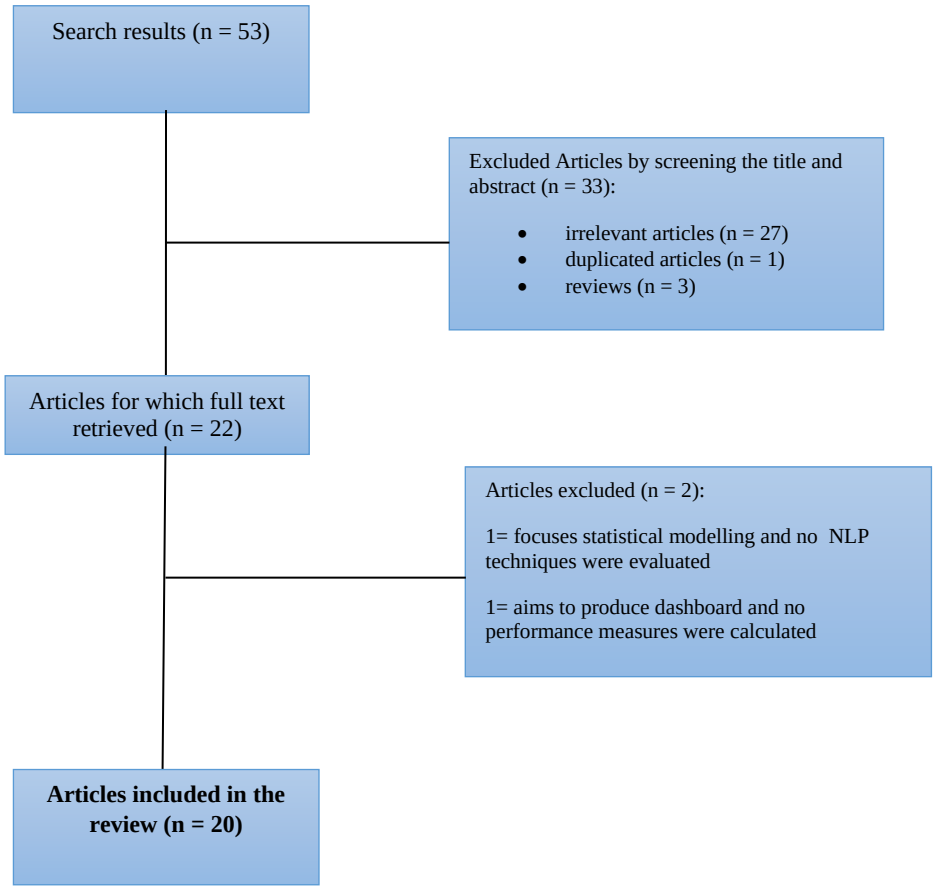


Figure 1: An example of preprocessing steps for text data using Natural Language Processing techniques

Figure (2): Flow diagram of the included studies in the review




Emerging Sources Citation Index (ESCI)
Clarivate
Analytics
WEB OF SCIENCE™
THOMSON
REUTERS