Supplemental Table 1. Detailed search strategy

Search terms	Search date	Database	Dates	Additional filters	Number of articles returned
1. ((natural language processing) OR (nlp)) AND ((cardiology) OR (cardiac) OR (cardiovascular))	4/10/2020	PubMed	1/1/2015 – 4/10/2020	English	200
2. ((natural language processing) OR (nlp)) AND ((cardiology) OR (cardiac) OR (cardiovascular))	4/10/2020	Embase	1/1/2015 – 4/10/2020	English	159
3. ((natural language processing) OR (nlp)) AND ((cardiology) OR (cardiac) OR (cardiovascular))	4/10/2020	Scopus	1/1/2015 – 4/10/2020	English	161
4. ((natural language processing) OR (nlp)) AND ((cardiology) OR (cardiac) OR (cardiovascular))	4/10/2020	ACM Digital Library	1/1/2015 – 4/10/2020	None	109
5. natural language processing) OR (nlp)) AND ((cardiology) OR (cardiac) OR (cardiovascular))	4/10/2020	IEEE Explore	1/1/2015 – 4/10/2020	None	20
6. natural language processing) OR (nlp)) AND ((cardiology) OR (cardiac) OR (cardiovascular))	4/10/2020	Arxiv	1/1/2015 – 4/10/2020	None	4
7. ((natural language processing) OR (nlp)) AND ((cardiology) OR (cardiac) OR (cardiovascular))	8/1/2021	PubMed	4/1/2020 – 12/31/2020	English	56
8. ((natural language processing) OR (nlp)) AND ((cardiology) OR (cardiac) OR (cardiovascular))	8/1/2021	Embase	4/1/2020 – 12/31/2020	English	67
9. ((natural language processing) OR (nlp)) AND ((cardiology) OR (cardiac) OR (cardiovascular))	8/1/2021	Scopus	4/1/2020 – 12/31/2020	English	52
10. ((natural language processing) OR (nlp)) AND ((cardiology) OR (cardiac) OR (cardiovascular))	8/1/2021	ACM Digital Library	4/1/2020 – 12/31/2020	None	23
11. natural language processing) OR (nlp)) AND ((cardiology) OR (cardiac) OR (cardiovascular))	8/1/2021	IEEE Explore	4/1/2020 – 12/31/2020	None	0
12. natural language processing) OR (nlp)) AND ((cardiology) OR (cardiac) OR (cardiovascular))	8/1/2021	Arxiv	4/1/2020 – 12/31/2020	None	0
13. (electronic health record* OR electronic medical record*) AND (information extraction OR named entity extraction OR named entity recognition OR coreference resolution OR relation extraction OR text mining) AND (cardiology OR cardiac OR cardiovascular) NOT (natural language processing OR nlp)	8/2/2021	PubMed	1/1/2015 – 12/31/2020	English	92

14. (electronic health record* OR electronic medical record*) AND (information extraction OR named entity extraction OR named entity recognition OR coreference resolution OR relation extraction OR text mining) AND (cardiology OR cardiac OR cardiovascular) NOT (natural language processing OR nlp)	8/2/2021	Embase	1/1/2015 – 12/31/2020	English	14
15. (electronic health record* OR electronic medical record*) AND (information extraction OR named entity extraction OR named entity recognition OR coreference resolution OR relation extraction OR text mining) AND (cardiology OR cardiac OR cardiovascular) NOT (natural language processing OR nlp)	8/2/2021	Scopus	1/1/2015 – 12/31/2020	English	12
16. (electronic health record* OR electronic medical record*) AND (information extraction OR named entity extraction OR named entity recognition OR coreference resolution OR relation extraction OR text mining) AND (cardiology OR cardiac OR cardiovascular) NOT (natural language processing OR nlp)	8/2/2021	ACM Digital Library	1/1/2015 – 12/31/2020	None	77
17. (electronic health record* OR electronic medical record*) AND (information extraction OR named entity extraction OR named entity recognition OR coreference resolution OR relation extraction OR text mining) AND (cardiology OR cardiac OR cardiovascular) NOT (natural language processing OR nlp)	8/2/2021	IEEE Explore	1/1/2015 – 12/31/2020	None	41
18. (electronic health record* OR electronic medical record*) AND (information extraction OR named entity extraction OR named entity recognition OR coreference resolution OR relation extraction OR text mining) AND (cardiology OR cardiac OR cardiovascular) NOT (natural language processing OR nlp)	8/2/2021	Arxiv	1/1/2015 – 12/31/2020	None	0

Supplemental Table 2. Detailed exclusion cascade			
	Round 1: April 2020	**Round 2: August 2021**	**Total**
Retrieved	654	434	1,088
Pubmed	200	148	348
Scopus	161	64	225
Embase	159	81	240
ACM Digital Library	109	100	209
IEEE Explore	21	41	62
Arxiv	4	0	4
Duplicates excluded	192	69	261
Title/ abstract: excluded	268	339	607
Non-cardiology focus	148	178	326
No NLP focus/details	87	94	181
Duplicate	31	19	50
Non-English	2	0	2
Review/perspective article	0	22	22
Abstract only	0	25	25
Outside of date range	0	1	1
Full-text: excluded	161	22	183
Non-cardiology focus	59	6	65
No NLP focus/details	42	10	52
Similar research study	13	0	13
Abstract only	35	4	39
Review/perspective article	12	0	12
Outside of date range	0	2	2
Records included	33	4	37

Supplemental Table 3. Patient populations, datasets, and NLP methods of included studies

	Patient population	Patient sample size	Patient demographic characteristics	Setting or dataset	No. of documents	Document types	NLP Tools and Methods	NLP Evaluation Methods
Adekkanattu et al, 2019[44]	Patients undergoing echocardiogram	600	Not reported	3 academic medical centers in the US and the MIMIC III database	600	Echocardiogram reports	Information extraction using *EchoExtractor* tool, implemented through *Leo* NLP system	Manual annotation by 2 clinicians and 1 researcher
Alnazzawi et al, 2016[29]	Heart failure patients	Not reported	Not reported	PhenoCHF corpus: discharge summaries (a subset of the documents from the i2b2 recognizing obesity challenge) and scientific articles	310	Discharge notes (n=300); scientific articles (n=10)	Named entity recognition with normalization methods (novel tool: *PhenoNorm*)	Previously annotated gold standard corpora linking entity mentions to other terminological resources
Bean et al, 2019[22]	Hospitalized AF patients	10,030	Mean age 75.3 (SD 12.3); 56.6% male	Single hospital in the United Kingdom	17,387	Discharge summaries	Named entity recognition (previously developed tool: *SemEHR*)	Manual annotation by two expert clinicians
Bielinski et al, 2015[30]	Heart Failure patients	110,110	MayoGC: mean age 65 (SD 12); Group Health: mean age 90 (SD 10); sex and race reported separately by HFpEF/HFrEF	eMERGE Cohort and multiple Mayo Clinic cohorts	110,110	Genomic data; inpatient and outpatient notes	Sectionizer to detect note sections and rule-based methods to identify concepts and assign status modifiers (positive, negative, probable) (previously developed tool: *MedTagger*)	Manual annotation by trained medical chart abstractors and nurse abstractors
Eggerth et al, 2020[42]	Heart failure patients	106	Mean age 71.1 (SD 12.1), 30% male	Austrian HF disease management network HerzMobil Tirol	3,952	Outpatient collaborative care notes	Classification: Bag-of-words model; tokenization: spaCy; classification: binary classifier trained using stochastic gradient descent learning with the machine-learning framework scikit-learn	Manual annotation (number of annotators not reported)
Esteban et al, 2017[15]	Cardiovascular and cerebrovascular disease patients	1,106	Not reported	Single hospital in Argentina	1,106	Inpatient, outpatient, emergency room notes	Named entity recognition and rule-based methods	Manual annotation by 3 family physicians
Evans et al, 2016[36]	Hospitalized HF patients	16,971	Not reported	Regional health system in the US	Not specified	Inpatient notes	Key term search	The entire risk prediction model

								(but not NLP alone) was compared with and without unstructured data.
Galper et al, 2018[51]	Patients receiving TAVR or Mitraclip	Not reported	Not reported	FDA's MAUDE system for post-market surveillance and Transcatheter Valve Therapy Registry	4,951	Registry reports, post-market reports	Named entity recognition, word embeddings (Novel tool: *Boomerang NLP*)	Manual annotation
Garvin et al, 2018[39]	Hospitalized HF patients	1,083	Not reported	8 VA medical centers	45,703	Inpatient notes	Key term search, rule-based and machine learning methods (novel tool: *CHIEF*)	Manual annotation by 2 independent reviewers; cardiologist resolved conflicts
Hu et al, 2016[19]	Hospitalized acute coronary syndrome patients	2,930	Mean age 62.3 (SD 12.1) years; 71% male	Single hospital in China	2,930	Inpatient admission notes	Rule-based methods and machine learning (conditional random fields)	Manual annotation by 3 physicians
Hu et al, 2019[27]	Patients who received CRT	990	Mean age 71.6 (SD 11.8), 78.1% male, 87.2% White	Partners Healthcare Research Patient Data Registry	Not specified	Inpatient notes	Word embeddings and bag-of-words models	Held-out testing set to evaluate entire machine learning algorithm (not NLP alone)
Jonnalagadda et al, 2017[40]	HFpEF patients	3,200	Not reported	Academic medical center in the US	1,934,640	Inpatient notes	Named entity recognition	Manual annotation by experienced clinical research coordinator
Kaspar et al, 2018[31]	HF patients	71,625	Age 26% <65 years; 28% 65-74 years, 50% >74 years; 59% male	Single tertiary care facility in Germany	71,625	Inpatient notes	Key term search	Manual annotation by 1 physician
Leiter et al, 2020[43]	HFrEF patients undergoing CRT	990	Mean age 71.2 (SD 12.2) years, 80.9% male, 86.1% White	Partners HealthCare Research Patient Data Registry	10,870	Discharge notes	Deep NLP algorithm for information extraction, Graph-IE, using local sequential and nonlocal coreferential dependencies between the words	Manual annotation by three study team members
Liu et al, 2019[37]	Hospitalized HF patients	Not reported	Not reported	MIMIC III database	13,746	Discharge notes	Machine learning methods	Basic random forest machine

							(convolutional neural networks)	learning algorithm
Mahajan et al, 2019[38]	Hospitalized HF patients	1,629	Not reported	6 VA medical centers	136,963	Inpatient notes	General information extraction methods (details not provided)	Held out validation set; compared to structured predictors alone and combined structured/ unstructured model
Moon et al, 2019[26]	Hypertrophic cardiomyopathy patients	200	Mean age 61, 55% male, 89% White	Mayo Hypertrophic Cardiomyopathy Registry	16,270	Inpatient notes	Rule-based methods and named entity recognition (previously developed tool: *MedTagger*)	Manual annotation by 2 reviewers, billing codes, patient surveys
Moon et al, 2020[25]	Hypertrophic cardiomyopathy patients	1,127	Not reported	Academic medical center in the US	687	ICD interrogation reports and electrophysiology notes	Rule-based methods (previously developed tool: *MedTagger*)	Manual annotation by 2 reviewers
Nath et al, 2016[45]	Cardiomyopathy patients	1,683	Mean age 67.9 (SD 13.9), 67.6% male	Academic medical center in the US	15,116	Echocardiogram reports	Rule-based methods to identify key terms and relationships between concepts (novel tool: *EchoInfer*)	Manual annotation by 2 reviewers, conflicts resolved by third reviewer
Owlia et al, 2019[16]	Patients with stable angina	6,556,919	Mean age 66.6 (SD 9.8) years; 99% male; 81% White	Veterans Health Administration (VA) clinical and administrative database	1,856,340	Inpatient and outpatient notes	Named entity recognition and rule-based methods (previously developed tool: *Leo*)	Chex validation tool and eHOST applications applied to held-out validation dataset
Patel et al, 2018[32]	HFpEF patients	80,248	Mean age 72 years; 96% male; 88% White	National VA healthcare database	Not specified	Inpatient and outpatient notes	Named entity recognition and rule-based methods (previously developed tool: *Leo*)	Manual annotation by 3 reviewers
Patterson et al, 2017[48]	HIV infected and uninfected patients undergoing echocardiogram	54,747	Not reported	National VA healthcare database	445,487	Outpatient notes, echocardiogram reports, radiology reports	Named entity recognition and rule-based methods (previously developed tool: *Leo*)	Manual annotation
Rosier et al, 2016[24]	AF patients	60	Not reported	Academic medical center in France	1,783	Inpatient and outpatient notes, lab reports, radiology reports	Key term search	Manual annotation by 2 physicians

Safarova et al, 2016[17]	Patients with elevated low-density lipoprotein (≥190)	6,547	Mean age 64.7 (SD 14.7) years; 45% male; 94% White	Mayo Employee and Community Health (ECH) system	6,547	Outpatient notes	Named entity recognition (previously developed tool: *MedTagger*)	Manual annotation by 1 physician
Shah et al, 2019[18]	Patients with myocardial infarction	2,000	Median age 75 (IQR 63-83) years; 61% male	Clinical Practice Research Datalink (CPRD), a population-based source of longitudinal clinical information in the United Kingdom.	31,913	Outpatient notes	Key term search and rule-based methods (novel tool: *Freetext Matching Algorithm*)	Manual annotation by 2 physicians
Shah et al, 2020a[23]	AF patients	786	Patients identified with AF: mean age 69.0 (SD 14.2) years, 60.6% male, 89.1% White	Academic medical center in the US	22,000	Inpatient and outpatient notes, lab reports, radiology reports	Key term search (previously developed tool: *pyConText*)	Manual annotation by 3 clinicians
Shah et al, 2020b [21]	AF patients	561	Site 1: Mean age 67.7 (SD 15.0); 43% female; 83-84% White Site 2: Mean age 70.0 (SD 15.1); 38% female; 61-84% White	Two academic medical centers in the US	1,800,000	Patient, visit, clinical, operational, financial, and research data from clinical data warehouses	Model built using logistic regression, extra trees, and naive Bayes classifiers	Manual annotation
Shi et al, 2015[46]	Congenital heart disease patients	Not reported	Not reported	Open-source forum associated with a hospital in China	3,464	Echocardiogram reports	Named entity recognition	Manual annotation by 5 reviewers
Toerper et al, 2016[20]	Patients undergoing cardiac catheterization	13,932	32% age 70 years and older; 64% male	Single urban hospital in the US	Not specified	Inpatient and outpatient notes	Key term search and rule-based methods	The entire forecasting algorithm (but not NLP alone) was validated using 10-fold cross-validation.
Topaz et al, 2017[41]	Hospitalized HF patients	8,901	Mean age 72.5 (SD 14.3) years; 57.2% male; 82.7% White	Academic medical center in the US	8,901	Discharge summaries	Named entity recognition, semantic analysis, and rule-based methods (previously developed tool: *MTERMS*)	Manual annotation by 2 reviewers
Valtchinov et al, 2020[49]	Patients with cardiac implantable devices	Not reported	Not reported	Academic medical center in the US	240,854	Inpatient and outpatient notes, radiology reports, procedure notes, microbiology reports	Expert-derived system, *QPID*, and ontology-derived rule-based methods (previously developed tool: *cTakes*)	Manual annotation by 1 researcher

Viani et al, 2019[28]	General cardiology patients	Not reported	Not reported	Molecular cardiology lab at a hospital in Italy	75	Notes describing clinical and family history, tests, medications, and diagnoses	Recurrent neural networks (RNNs)	Manual annotation by 2 researchers
Wagholikar et al, 2018[33]	HFrEF patients	57,158	Not reported	Academic medical center in the US	46,634	Echocardiogram reports	Key term search	Manual annotation by clinical experts
Wang et al, 2015[34]	Hospitalized HF patients	18,295	Not reported	Health Information Exchange in the US	2,139,299	Inpatient and outpatient notes	Machine learning methods (random forest/decision tree models)	Manual annotations by 2 physicians
Xie et al, 2017[47]	Patients undergoing echocardiogram	621,856	51.2% age $\geq$ 65; 49.5% male	Large regional health system in the US	621,856	Echocardiogram reports	Key term search and normalization methods	Manual annotation by 1 cardiologist
Zhang et al, 2018[35]	HF patients with a CRT device	36,276	Not reported	Academic medical center in the US	6,174	Inpatient and outpatient notes	Rule-based and multiple machine-learning methods (bag-of-words, support vector machine, logistical regression, and random forest)	Structured documentation of NYHA class
Zheng et al, 2020[50]	Patients presenting to the emergency department who had a troponin laboratory test and underwent an ETT within 30 days of their ED visits	5,214	Mean age 56 years, 49.6% male, 48.1% white	Large regional health system in the US	5,214	Exercise treadmill test reports	Rule-based methods to identify key terms and relationships between concepts	Manual review by one emergency physician and one cardiologist

AF: atrial fibrillation; CRT: cardiac resynchronization therapy; EHR: electronic health record; HF: Heart failure; HFrEF: Heart failure with reduced ejection fraction; HFpEF: Heart failure with preserved ejection fraction; NLP: natural language processing

Supplemental Table 4. Purpose and findings of included studies by cardiac disease focus

	Study purpose	Performance of NLP algorithm	Study outcomes
Coronary Artery Disease			
Esteban et al, 2017[15]	To assess the sensitivity, specificity, and agreement level of a rule-based algorithm for the detection of cardiovascular and cerebrovascular disease.	Sensitivity: 96-99% Specificity: 86-97%	The developed algorithm used only standardized and non-standardized coded terms within an EHR to properly detect clinically relevant events and symptoms of cardiovascular and cerebrovascular disease.
Hu et al, 2016[19]	To explore major adverse cardiovascular event (MACE) prediction in a proactive manner using inpatient admission records.	AUC: 72%	NLP algorithm predicted MACE for acute coronary syndrome patients at the early stage of their hospitalizations significantly better than two well-known acute coronary syndrome risk score tools.
Owlia et al, 2019[16]	To use NLP to extract Canadian Cardiovascular Society (CCS) angina severity classifications from clinical notes and determine associations between CCS and all-cause mortality and healthcare utilization.	PPV: 93% Sensitivity: 76%	NLP–extracted CCS classification was positively associated with all-cause mortality and healthcare utilization, demonstrating the prognostic importance of anginal symptom assessment and documentation.
Safarova et al, 2016[17]	To develop an ePhenotyping algorithm for rapid identification of familial hypercholesterolemia.	PPV and NPV $\geq$ 85%	The algorithm identified patient with familial hypercholesterolemia, many of whom had not been previously identified.
Shah et al, 2019[18]	To describe the contribution of outpatient notes in the 90 days prior to a myocardial infarction to the recording of information about myocardial infarction (subtype, left ventricular function, laboratory results and symptoms) and cause of death.	PPV: 83-92% Sensitivity: 17-41% Specificity: 96-97%	Outpatient notes contained information such as symptoms, results and specific diagnoses that could be useful in characterizing a myocardial infarction.
Toerper et al, 2016[20]	To develop and prospectively evaluate a web-based tool that forecasts the daily bed need for admissions from the cardiac catheterization laboratory using structured and unstructured EHR data.	AUC: 72%	The forecast model identified older age, male gender, invasive procedures, coronary artery bypass grafts, and a history of congestive heart failure as qualities indicating a patient was at increased risk for admission following catheterization.
Electrophysiology			
Bean et al, 2019[22]	To develop and validate an NLP risk scoring pipeline, explore trends in antithrombotic medication use for AF, and quantify the association between antithrombotic medication use and relevant clinical patient-level variables.	PPV: 95% Sensitivity: 97% F-score: 96% Accuracy: 98%	Automatic risk scores were in strong agreement with the two independent experts for CHA2DS2-VASc. Agreement was lower for HAS-BLED.
Hu et al, 2019[27]	To apply machine learning to create an algorithm that predicts CRT outcomes using EHR data.	PPV: 79% Sensitivity: 26% F-score: 77% Accuracy: 65% AUC: 75%	A machine learning model that leveraged readily available EHR data and clinical notes identified a subset of CRT patients who may not benefit from CRT before the procedure.
Moon et al, 2019[26]	To develop and deploy NLP algorithms for automated extraction of syncope, family history of sudden cardiac death, and family history of hypertrophic cardiomyopathy from clinical narratives.	NPV: 90-97% Sensitivity: 91-95% Specificity: 90-98%	Automated extraction of the desired data elements using NLP is feasible and has promise to increase efficiency of workflow for providers managing hypertrophic cardiomyopathy patients.
Moon et al, 2020[25]	To compare the performance of unstructured notes to structured device data in determining heart rhythm from implantable cardioverter defibrillators (ICDs) and whether patients with hypertrophic cardiomyopathy received appropriate therapy from their ICD.	NLP only model F-score: 0.92-0.98 Structured data only model F-score: 0.45-0.78	The NLP methods on unstructured data performed significantly better in identifying rhythm and therapy delivered from ICDs than the methods using structured data only.
Rosier et al, 2016[24]	To design a prototype mechanism for AF alerts in remote monitoring of CIEDs, and to evaluate the efficacy and safety of this prototype.	Accuracy: 98%	The alert classification system including the NLP classifier had high agreement (kappa= 0.93) with the manual annotation of alerts.
Shah et al, 2020a[23]	To compare the samples and characteristics from each model, and oral anticoagulant treatment rates in each sample, of various models using structured and unstructured data to identify AF patients in EHRs.	NLP only model Sensitivity: 97% Combined NLP/ structured data model Sensitivity: 90% Specificity: 87%	Models that included unstructured combined with structured data were the most accurate at identifying patients with AF.

		Specificity: 63% AUC: 80%	AUC: 89%	
Shah et al, 2020b[21]	To create a portable NLP algorithm to identify patients with atrial fibrillation (AF) using text alone.	Sensitivity: 90-93% Specificity: 71-89% AUC: 80-91% F-score: 93-94%		The NLP algorithm was able to identify patients with AF using text alone with >90% F-score at 2 separate sites, creating opportunities for precise, high-throughput cohort identification.
General cardiology				
Viani et al, 2019[28]	To apply novel recurrent neural networks (RNNs) for event extraction from medical reports in the cardiology domain written in Italian.	RNN classifier alone: PPV: 88% Sensitivity: 89% RNN and dictionary lookup: PPV: 87% Sensitivity: 92% F-score: 90%		Integrating a well-performing RNN-based classifier with a standard knowledge-based approach can be a good strategy to extract information from clinical text in non-English languages.
Heart Failure				
Alnazzawi et al, 2016[29]	To develop a novel method, PhenoNorm, which integrates similarity measures to allow automatic linking of phenotype concept mentions to known concepts in the UMLS Metathesaurus, a biomedical terminological resource.	F-score: 76-83% Accuracy: 77-86%		PhenoNorm outperforms a number of alternative methods applied to the same task and has wider utility.
Bielinski et al, 2015[30]	To develop and validate an EHR-based algorithm to accurately identify HF patients with characterization of HFpEF and HFrEF.	PPV: 80-94% NPV: 98-100% Sensitivity: 71-100% Specificity: 97-99%		The developed algorithm was expanded to include definite, probable, and possible HF based on the degree of confidence of the classification to capture HF cases.
Eggerth et al, 2020[42]	To develop classifiers for automated categorization of collaboration notes documenting medication management using NLP.	PPV: 75-99% Sensitivity: 44-82% F-score: 55-90%		The classifier achieved high accuracy and could be used to evaluate medication adherence and other aspects of management among HF patients.
Evans et al, 2016[36]	To develop and evaluate an automated identification and predictive hospitalized HF patients' risk of 30-day readmissions and mortality.	PPV: 98% Sensitivity: 95% Specificity: 98%		The addition of NLP-identified HF patients significantly improved model performance in identification and prediction activities.
Garvin et al, 2018[39]	To accurately automate a United States Department of Veterans Affairs quality measure for inpatients with HF.	PPV: 89-99% Sensitivity: 27-100% F-score: 42-100%		The NLP methods provided accurately classified patients with HF with high performance for meeting care metrics.
Jonnalagadda et al, 2017[40]	To develop a high recall prescreening algorithm for recruiting patients into a multicenter, randomized controlled study.	PPV: 86% Sensitivity: 95%		Automated identification of HFpEF patients who are appropriate candidates for a clinical trial is feasible and time efficient.
Kaspar et al, 2018[31]	To approximate the "true number" of patients suffering from HF at a tertiary care center.	PPV: 71-96% Sensitivity: 60-92% F-score: 74-86%		The NLP-based search algorithm markedly improved diagnostic accuracy compared to ICD codes alone.
Leiter et al, 2020[43]	To develop, train, and evaluate a deep NLP algorithm to identify documented symptoms from unprocessed EHR notes in a cohort of patients with HFrEF who would subsequently undergo CRT.	PPV: 78% Sensitivity: 67% Accuracy: 99% F-score: 72%		The deep NLP algorithm trained to capture symptoms in patients with CHF who received CRT showed promising precision (PPV) and recall.
Liu et al, 2019[37]	To evaluate a deep learning approach to predict heart failure readmission from clinical notes.	F-score: 73-76%		Deep learning models outperformed the regular models in prediction tasks.
Mahajan et al, 2019[38]	To use machine learning methods such as statistical NLP for predicting the risk of readmission for heart failure.	AUC: 51-65%		Predictive models using both structured and unstructured EHR data best predicted HF readmissions compared to models using only structured and only unstructured data.
Patel et al, 2018[32]	To use NLP extraction of LVEF values and algorithm development and validation to derive an algorithm that is able to curate a cohort of HFpEF from a large national database.	PPV: 96% NPV: 87% Sensitivity: 88%		An algorithm using structured and unstructured EHR data led to the creation of a HFpEF cohort within the health system.

		Specificity: 96%	
Topaz et al, 2017[41]	To identify HF patients with ineffective self-management by using NLP on a subsample of discharge.	PPV: 95% Sensitivity: 79%	NLP successfully identified HF self-management behaviors in discharge notes. Specific types of self-management deficits were significantly associated with readmissions.
Wagholikar et al, 2018[33]	To use a regular expression-based NLP system to extract LVEF from echocardiogram reports to identify patients presenting with HFrEF for driving a population-based therapeutic intervention program.	Accuracy: 100%	The regular expression-based approach accurately extracted LVEF from echocardiograms and was useful for identifying HFrEF patients.
Wang et al, 2015[34]	To identify HF cases from both EHR codified and NLP found cases.	PPV: 91%	A HF case finding algorithm including NLP was developed, tested, prospectively validated, and integrated into the Health Information Exchange live system.
Zhang et al, 2018[35]	To extract NYHA class for patients with a CRT device from electronic health records.	Rule-based methods PPV: 95% Sensitivity: 92% F-score: 94% Machine-learning methods PPV: 85-95% Sensitivity: 81-93% F-score: 83-94%	Machine learning-based methods outperformed a rule-based method. The best machine-learning method was a random forest with n-gram features.
Imaging			
Adekkanattu et al, 2019[44]	To investigate the portability of an NLP system to extract 27 key cardiac concepts from echocardiogram reports.	PPV: 6-100% Sensitivity: 25-100% F-score: 11-100	The performance of the NLP methods varied widely based on the cardiac concept of interest.
Nath et al, 2016[45]	To explore the feasibility and reliability of using NLP for large-scale and targeted extraction of multiple data elements from echocardiogram reports.	PPV: 94% Sensitivity: 92% F-score: 93	The NLP methods reliably extracted relevant data elements from echocardiography reports.
Patterson et al, 2017[48]	To develop a robust and efficient clinical information extraction system utilizing both coded and unstructured data towards the goal of refining heart failure phenotypes.	PPV: 94-98% Sensitivity: 74-82%	The NLP methods that were developed feasibly and effectively extracted HF-related information.
Shi et al, 2015[46]	To extract anatomic site-related features in echocardiogram reports and predict risk level.	Rule-based methods PPV: 51% Sensitivity: 43% F-score: 44 Machine learning methods PPV: 47% Sensitivity: 42% F-score: 43 AUC: 61%	The machine learning methods slightly outperformed rule-based methods, but performance varied by risk level.
Valtchinov et al, 2020[49]	To assess two NLP approaches (expert-derived and ontology-derived) to identify patients with implantable devices that pose safety risks for MRI.	Expert-derived NLP Sensitivity: 88% Specificity: 82% Accuracy: 83% Ontology-derived NLP Sensitivity: 96% Specificity: 92% Accuracy: 91%	Both expert- and ontology-derived NLP approaches have similar accuracy in identifying patients with implantable devices that pose high safety risks for MRI.
Xie et al, 2017[47]	To develop a computerized NLP algorithm to extract EF from echocardiogram reports.	PPV: 97% Sensitivity: 95%	The NLP algorithm achieved high performance.
Zheng et al, 2020[50]	To develop and validate an automated method to interpret exercise treadmill test results and test the associations between exercise treadmill test results with 30-day patient outcomes in a large population.	Sensitivity: 96% Specificity: 95%	NLP effectively extracted relevant information from exercise treadmill tests. Most patients were at low risk of severe 30-day outcomes (myocardial infarction, death) and had normal ETT results.
Valvular Disease			

Galper et al, 2018[51]	To analyze reported adverse events and device problems for both TAVR and Mitraclip and compared rates between unstructured data in FDA's MAUDE system for post-market surveillance and structured data the Transcatheter Valve Therapy Registry.	<u>NLP methods (vs structured data)</u> R^2: 0.86	NLP enabled identification of the most common events associated with TAVR and Mitraclip procedures. Event rates derived from the two data sources were not significantly different.
AF: atrial fibrillation; CRT: cardiac resynchronization therapy; EHR: electronic health record; HF: Heart failure; HFrEF: Heart failure with reduced ejection fraction; HFpEF: Heart failure with preserved ejection fraction; NLP: natural language processing			