Engineering and Technology Journal e-ISSN: 2456-3358

Volume 10 Issue 05 May-2025, Page No.- 5227-5239

DOI: 10.47191/etj/v10i05.51, I.F. – 8.482

A Review of Natural Language Processing for Structured and Unstructured Data in Electronic Health Records

Kazheen Ismael Hasan[1], Ibrahim Mahmood Ibrahim[2]

[1]Information Technology Department, Technical College of Informatics-Akre, Akre University for Applied Sciences, Duhok, KRG - Iraq,

[2]Computer Networks and Information Security Department, Technical College of Informatics-Akre, Akre University for Applied Sciences, Duhok, KRG - Iraq

ABSTRACT: This study rigorously analyses the development, methodologies, and efficacy of natural language processing (NLP) applications in healthcare literature from 2020 to 2025. A systematic review of 23 peer-reviewed articles examines how methodological diversity— ranging from rule-based systems to deep learning and large language models—has impacted clinical insights and research outcomes. The analysis indicates a dual trajectory: advanced models like BERT, EXGB, and LSTMs excel with high-performance metrics (F1 $\geq$ 0.85 in 41% of studies), while traditional rule-based methods endure due to their transparency, specificity, and practicality in low-resource environments. A significant limitation persists: more than 25% of the studies either lack quantitative assessment or depend exclusively on qualitative illustrations, compromising reproducibility and generalizability. Moreover, the limited range of datasets—typically confined to single institutions or specific tasks elicits apprehensions regarding model robustness and external validity. This paper contends that the field is at a pivotal juncture: future advancements require technical expertise and a transition towards interdisciplinary collaboration, standardization in assessment, and ethical awareness in applying NLP in clinical settings. We propose a progressive research agenda harmonizing innovation with accountability and contextual awareness.

KEYWORDS: Natural Language Processing (NLP), Structured Data, Unstructured Data, and Electronic Health Records (HER).

1. INTRODUCTION

Leveraging structured and unstructured data, Natural Language Processing (NLP) has evolved into a necessary instrument in EHR management and analysis, improving patient care and health outcomes [1]. Electronic health records have transformed healthcare by building a huge database of organized fields like checkboxes and coded entries, and unstructured fields like clinical notes offer qualitative patient health insights [2]. According to the literature, NLP applications for clinical and research uses are developing quickly to extract pertinent information from many data formats [3].

Extensively clinically relevant insights from unstructured textual data in EHRs are increasingly extracted from NLP technique [4].demonstrated how NLP might easily combine quantitative assessments with qualitative data extracts from clinician notes to ascertain the Expanded Disability Status Scale score from patient records[5]. According automated NLP solutions could spot symptom onset dates in clinical notes, so offering time-sensitive information for cardiology patient management and intervention plans [6].These uses show how NLP enhances temporal integration and data extraction, thus improving healthcare delivery and outcomes [7].

Research on NLP techniques mirrors the increasing expenditure in this are[8]highlighted algorithm design innovations that enhance unstructured data extraction and analysis [9]outlining a complete NLP pipeline for chronic lupus disease[10]. As NLP's ability to forecast diabetic patient risks[11]shows, these developments are relevant outside of specific circumstances. More particular illustrations or direct references would support this assertion.

Rich, multifarious data from EHRs presents challenges in processing [12]. Though progress has been achieved, switching from conventional data management to advanced NLP systems requires a grasp of complex clinical narratives [13] pointed out that NLP can derive thorough clinical phenotypes from unstructured narratives [14], thus allowing a complete knowledge of schizophrenia cognitive deficits [15] This is a big step since conventional structured data usually misses the complexity of mental health diagnosis[16].

how automated approaches can expose contextual information in clinical narratives[17], including exposure to gun violence, stressing the need for NLP approaches that fit evolving healthcare and social concerns [18] Data standardization, ethical issues, and integration with healthcare systems make NLP technology research and development especially important[19]. As researchers push NLP's boundaries in electronic health record analysis, the literature demonstrates a great will to solve these difficulties [20].

In healthcare, natural language processing has creative uses to enhance clinical judgment [21]. From free-text notes[12], systems have been shown to extract and classify medical conditions, drugs, and laboratory results [22] Using textual data, NLP techniques examined cardiac care delays to find elements influencing patient outcomes and treatment schedules [17]This capacity to derive practical insights from unstructured data enables doctors to grasp healthcare system inefficiencies and guide their decisions[23].

Deep learning and machine learning paradigms have enhanced NLP models' ability to grasp the idiomatic and contextual character of medical language. These developments let NLP forecast clinical outcomes and enhance patient-clinician contact [24]EHR text toolkit emphasizes the need for user-friendly frameworks for NLP technology adoption across clinical environments [24]. By democratizing advanced data analytics, these developments let many healthcare facilities use NLP for better patient care[25]

Finally, in electronic health records, natural language processing changes analysis and management of healthcare data. Recent studies reveal that NLP can turn unstructured clinical narratives into ordered[26], useful insights to change patient care. As the field develops, data standardization, interoperability[27], [28], and ethical issues must be carefully considered to maximize the clinical use of NLP technologies. Therefore, technologists, medical professionals, and policymakers must work together to maximize NLP's potential in healthcare systems [29].

2 BACKGROUND THEORY

2.1 Structure and Complexity of EHR Data

Essential tools for clinical documentation and data-driven healthcare, Electronic Health Records (EHRs) retain a vast spectrum of patient data[29] One can categories EHR data into structured and unstructured forms. Easily stored, queried, and examined using conventional techniques, structured data consists of predefined fields, including ICD codes,[30] lab results, and medication lists. Unstructured data, on the other hand, comprises physician notes, radiology reports, and clinical narratives—rich in detail but difficult to evaluate because of their variability and lack of standardizing. Comprehensive patient knowledge depends on these free-text components, often including insightful clinical background not found in structured entries[31].

2.2 Evolution of NLP in Healthcare

Natural language processing (NLP) is an artificial intelligence subfield that lets machines understand and use human language. In the context of healthcare, NLP has developed rather dramatically over time[32]. Early systems extracted data from text using dictionaries and manually defined patterns, following rules. For example[33]showed how a rule-based pipeline using IBM Watson Content Analytics could precisely identify terms linked to epilepsy from unstructured clinical letters[34]. These systems lacked scalability across many datasets and clinical scenarios but were highly interpretable and efficient in limited use cases[35]. More adaptive approaches were adopted as clinical documentation grew in complexity and volume since rule-based systems proved inadequate [36].

2.3 Advances in Machine Learning and Deep Learning

Machine learning and deep learning represent a sea change in clinical natural language processing. Learning from labelled data lets supervised learning methods automatically extract pertinent clinical features[37], including logistic regression and support vector machines (SVM). More recently, deep learning models, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformers like BERT, have successfully captured complicated contextual relationships in free-text data. Research, including those by [38],[39], shows how these models are now applied to forecast outcomes and extract disease-specific data from unstructured notes. These developments have greatly raised NLP applications in EHRs' scalability, adaptability, and accuracy[40].

2.4 Literature Review

Though much has been done, several obstacles still limit NLP's complete integration into clinical practice. Inconsistent terminology, strong use of abbreviations, and variation in documentation styles between institutions all lower model generalizability. Many NLP models, particularly deep learning-based ones, behave as "black boxes," making their predictions challenging to understand a major issue in clinical decision-making. Most NLP tools also depend on annotated data, which is sometimes limited and expensive, especially for rare diseases or sensitive medical conditions. As[31] have pointed out, even modern models need careful tuning and evaluation before clinical release. Future studies must concentrate on creating scalable, generalizable, interpretable NLP systems that easily merge into clinical processes.

Rajendran and Topaloglu (2020)[41] built an NLP and deep learning pipeline to categories smoking status from clinical notes. They found that CNN with pre-trained embeddings attained the best performance: 80.7% accuracy and an 85.4% F1 score, using CNNs and LSTMs with pre-trained and custom word embeddings. Their deep learning models outperformed conventional techniques, including Naive Bayes and SVM, highlighting the promise of automated NLP for extracting insightful clinical information from unstructured data.

Solomon et al. (2021)[42] Examining over 927,000 reports, their model produced predictive values above 95%, while just 59% with diagnosis codes achieved. The NLP method also found several notable cases lost by administrative data. The paper emphasizes how NLP might enhance tailored treatment and disease monitoring. It shows how well NLP pulls clinical conditions from massive unstructured data.

Lo Barco et al. (2021) [43]employed natural language processing to facilitate the early diagnosis of Dravet syndrome by examining unstructured electronic health record reports. They conducted a comparison of DS and febrile seizure cases utilizing UMLS-based concept extraction. DS reports included a greater frequency of terms such as "prolonged seizures," "myoclonia," and "regression," which were infrequently observed in febrile seizures. The research underscores NLP's capability to identify rare diseases early by revealing nuanced phenotypic patterns. This method may diminish diagnostic delays and enhance patient outcomes.

Tsui et al. (2021)[44] employed natural language processing and machine learning to forecast initial suicide attempts by examining both structured and unstructured electronic health record data. Integrating both data types enhanced prediction accuracy (AUC 0.932 compared to 0.901). The leading model, extreme gradient boosting, utilised over 1,700 features and exhibited strong performance across multiple timeframes. NLP identified critical risk factors absent in structured data, including social determinants and previous attempts. The research underscores the efficacy of combined NLP-ML methodologies for the early identification of suicide risk.

Hatef et al. (2022) [45]devised a rule-based NLP algorithm to detect residential instability, such as homelessness, from unstructured clinical notes within three healthcare systems. They enhanced the algorithm with site-specific terminology utilizing spaCy tools and expert-curated patterns. Performance exhibited variability across sites, with precision fluctuating between 0.45 and 1.00 and sensitivity ranging from 0.68 to 0.96, contingent upon documentation quality. The research illustrates that rule-based natural language processing can proficiently identify social determinants of health when customized for institutions. It underscores the necessity for customization while pursuing extensive applicability.

Han et al. (2022)[46] created a deep learning-based natural language processing framework to categorize 13 social determinants of health from unstructured electronic health record notes in the MIMIC-III dataset. They trained CNN, LSTM, and BERT models, with BERT attaining the highest performance (micro-F1 = 0.690, macro-AUC = 0.907). BERT demonstrated superior proficiency in categorizing "occupational" and "support circumstances," surpassing conventional models. The research underscores the significance of transformer-based models in augmenting structured data with social context derived from clinical narratives.

Henrique et al. (2022)[47]created an automated NLP system to extract structured data from unstructured Portuguese electronic patient records. The system employed tokenization, morphological analysis, and Levenshtein distance to analyze clinical notes from pregnant women in a Brazilian maternity hospital. A medical dictionary was developed with expert contributions to categorize 12 types of clinical information. The system's performance is closely aligned with manual annotations, exhibiting no substantial discrepancies in accuracy. This study emphasizes the capabilities of domain-specific NLP tools in non-English healthcare environments.

Lybarger et al. (2023) [48]created mSpERT, a deep learning model to extract social determinants of health (SDOH) from clinical narratives. mSpERT, trained on the Social History Annotated Corpus, attained an F1 score of 0.86. Upon analysis of over 430,000 social history notes, it was determined that 32% of homelessness cases and 19% of tobacco use instances were recorded exclusively in unstructured text. This stance underscores the model's ability to augment structured EHRs with essential social risk data.

Liu et al. (2023) [49]devised a natural language processing method to ascertain lung cancer screening eligibility by integrating structured and unstructured smoking data from electronic health records. They optimized a rule-based instrument to extract metrics such as pack years and cessation dates, attaining an F1 score of 0.909 and 96% accuracy. This method identified 73.8% more eligible patients than structured data alone, with notable enhancements among Black/African American individuals. The research illustrates how natural language processing can enhance screening precision and mitigate healthcare inequalities.

Barman et al. (2023) [50]employed augmented curation NLP models to extract clinical data regarding propionic acidemia (PA) from electronic health records at the Mayo Clinic. Their BERT-based pipeline, validated by clinicians, detected 13 cases of PA, revealing six that were overlooked by structured data. The study demonstrated that patients experiencing metabolic decompensation events encountered a greater number of complications, underscoring the importance of comprehensive data extraction. This study illustrates how transformer-based natural language processing can enhance the comprehension of rare diseases in clinical environments.

Lin et al. (2024)[51] created a deep learning-based Named Entity Recognition system utilizing FinBERT to identify underreported health conditions such as incontinence, falls, loneliness, and mobility challenges from unstructured Electronic Health Records. The model, trained on over 10 million texts, surpassed ICD-10 codes by identifying more cases and demonstrating superior predictive value for all-cause mortality. It also disclosed an age-associated increase in these conditions. This study illustrates the efficacy of domain-specific NLP tools in enhancing healthcare insights and outcomes.

More (2024)[52] investigated the application of NLP to derive structured data from unstructured clinical notes in EHRs, tackling issues such as free-text variability and semantic ambiguity. A multi-step NLP pipeline was created, integrating named entity recognition, relation extraction, temporal analysis, and negation detection. The methodology effectively extracted medication information, including dosage and administration routes, utilizing SpaCy and the Med7 model. The research underscores the capability of lightweight, domain-specific NLP tools to convert clinical narratives into structured, actionable data for the healthcare sector.

Alskaf et al. (2024)[38] created an AI-based pipeline that employs natural language processing and machine learning to examine unstructured electronic health records and structured patient care management reports. Utilising tools such as CogStack and MedCAT, their NLP models attained substantial accuracy (F1 > 0.87) in extracting cardiovascular risk factors. The extracted data were utilised to train machine learning models for predicting all-cause mortality, with support vector machines exhibiting the highest performance (AUC = 0.82). The incorporation of imaging findings, including ischaemia and late gadolinium enhancement (LGE), with clinical data enhanced predictive accuracy. This study demonstrates the clinical significance of fully automated systems for processing and modelling unstructured cardiovascular health data.

Badalotti et al. (2024) [53]created a BERT-based NLP model to extract comorbidity data from unstructured Italian stroke patient electronic health records, utilizing active learning to improve annotation efficiency. Employing Bayesian uncertainty sampling (BALD), the model attained consistent performance (macro F1 > 0.7) after labelling 50% of the dataset. Prognostic models developed using semi-automated data corresponded with those created from manual annotations. The research underscores difficulties in recognizing rare comorbidities and advocates for AL-enhanced BERT models for clinical data extraction in non-English contexts.

Martina et al. (2024)[54] devised rule-based NLP algorithms to extract four social determinants of health (SDOH) from inpatient electronic medical records in Calgary. The model, validated across over 17,000 admissions, demonstrated high positive predictive values for language barriers (99%), living alone (98%), and unemployment (96%). The research indicated prevalent inaccuracies in structured EMR fields, with more SDOH recorded in unstructured notes. It underscores the significance of lightweight NLP in enhancing the comprehensiveness and clinical utility of EMRs.

Ruckdeschel et al. (2025)[55] illustrated that natural language processing applied to unstructured clinical notes more effectively captures smoking history than structured electronic health record data. Utilising the MIMIC-III dataset alongside a combination of rule-based and transformer models, they extracted information such as pack-years and cessation dates. Their methodology identified 41.8% of patients as former smokers, in contrast to 12.5% utilizing ICD-9 codes. The NLP model attained an F1 score of 0.88 in identifying patients eligible for screening, underscoring NLP's contribution to risk-based screening and personalized care.

Clay et al. (2025) [56]provided an extensive overview of NLP techniques for electronic health records, encompassing statistical and neural network methodologies. The document outlined the comprehensive NLP pipeline, encompassing preprocessing to sophisticated models such as BERT and GPT. It emphasized critical applications, including information extraction and patient classification. The authors highlighted the potential of large language models in clinical environments while acknowledging challenges such as explainability and bias. Their objective was to link clinical researchers with NLP tools to promote wider and more effective applications in healthcare.

Chen et al. (2025) [57]performed a bibliometric analysis of 4,803 articles related to NLP from 2011 to 2024 utilizing Cite Space and VOS viewer. They recognized significant trends and prominent contributors, observing an increase in NLP publications post-2018, predominantly from the U.S. and China. Principal research domains encompassed deep learning, electronic health records, sentiment analysis, and risk evaluation. The study emphasizes the increasing significance of NLP in interdisciplinary domains, especially in healthcare, providing an overview of the developing research landscape.

Takeuchi et al. (2025)[58] established a sequential NLP framework to extract tumor response, progression dates, and survival curves from unstructured electronic health records of lung cancer patients. They integrated rule-based methodologies with LSTM models to evaluate radiology reports and categorize tumor responses. The model produced survival curves that closely aligned with physician-reviewed data, demonstrating its utility for survival analysis. The research highlights the promise of interpretable NLP tools in oncology for developing significant clinical databases and obtaining longitudinal outcomes.

Bryde Alnor et al. (2025)[59] engaged natural language processing to examine electronic health records and discern significant bleeding incidents and associated risk factors in hospitalized patients. Their validated NLP model demonstrated that merely 39% of bleeding events were identified by ICD-10 codes, underscoring NLP's enhanced detection proficiency. Two Cox regression-based risk assessment models were developed, identifying significant predictors such as low hemoglobin, elevated creatinine, and alcohol consumption. Both models exhibited strong performance (C-statistics >0.72), with one model specifically tailored for real-time clinical application. The research illustrates the advantages of integrating NLP with laboratory data for enhanced risk assessment and surveillance.

Soto Jacome et al. (2024) [60]devised a rule-based NLP algorithm to evaluate the appropriateness of thyroid ultrasound (TUS) orders based on unstructured electronic health record (EHR) notes. The model attained elevated accuracy (sensitivity = 0.96, specificity = 0.80, PPV = 0.99) in distinguishing appropriate from inappropriate TUS orders. They also developed a condensed version utilizing requisition notes with comparable efficacy. The research underscores the capability of NLP for real-time appropriateness assessment in thyroid cancer management.

Li et al. (2025)[24] presented Adaptive-Mixture-Categorization (AMC), a preprocessing technique to enhance suicide risk prediction through NLP-derived data from VA electronic health records. AMC converted biassed data into resilient categories, surpassing raw and quantile-based approaches. Utilising 517 predictors from SÉANCE, AMC demonstrated superior efficacy in detecting suicide risk. The research underscores sophisticated preprocessing to improve NLP models for mental health forecasting.

Pan et al. (2025)[39] developed a pipeline combining large language models (LLMs) with clinical expert knowledge to detect conditions like AMI, diabetes, and hypertension from unstructured EHR notes. The Mistral-7B-OpenOrca system attained a sensitivity of 88-94%, exceeding ICD-10 coding in real-time surveillance. The pipeline incorporated prompt-based preprocessing, reasoning, and rule-based postprocessing to ensure clinical relevance. The research exhibited improved explainability and scalability of large language models for automated, interpretable phenotyping in healthcare.

3. Table 1. Comparison among the reviewed works

Table 1 provides a detailed summary of the previous publications addressed in Section 3. It presents the key indicators used for assessment and emphasizes the significant findings from these studies, highlighting the strengths and innovative concepts that emerged from the research.

Author & Year	Task	NLP Method	Data	Model	Performance	Dataset	Limitation
[41]Rajendran & Topaloglu (2020)	Smoking classification	DL (CNN, LSTM)	Unstructured	CNN (pretrained)	F1: 0.85, Acc: 0.81	6,298 notes	Small dataset limited performance
[42]Solomon et al. (2021)	Aortic stenosis detection	Rule-based	Semi-structured	I2E	PPV: 99%	927,884 echos	Single health system
[43]Lo Barco et al. (2021)	Dravet syndrome detection	Rule-based + UMLS	Unstructured	DrWarehouse	————	104 patients	WSD issues, generalizability
[44]Tsui et al. (2021)	Suicide risk prediction	ML + NLP (cTAKES)	Structured + Unstructured	EXGB	AUC: 0.93	45,238 patients	Single-site, static model
[45]Hatef et al. (2022)	Residential instability detection	Rule-based	Unstructured	spaCy	PPV: 0.45-1.0	3 health systems	Site-specific limitations
[46]Han et al. (2022)	SDOH classification	DL (BERT, CNN, LSTM)	Unstructured	BERT	F1: 0.69, AUC: 0.91	MIMIC-III	Limited to Social Work notes
[47]Henrique et al. (2022)	Structuring pregnancy records	Rule-based + Levenshtein	Unstructured	Java pipeline	Accuracy ~89%	4,000 EPRs	Single hospital, clinical exam extraction poor
[48]Lybarger et al. (2023)	SDOH extraction	DL (BERT)	Structured + Unstructured	mSpERT	F1: 0.86	430K notes	Limited categories,1-year data
[49]Liu et al. (2023)	LCS eligibility detection	Rule-based	Structured + Unstructured	Custom rules	F1: 0.91	102,475 patients	Clean input assumed

[50]Barman et al. (2023)	Propionic acidemia curation	DL (BERT)	Structured + Unstructured	Augmented Curation	F1: 0.91	13 PA cases	Tiny sample, retrospective
[51]Lin et al. (2024)	Aging condition NER	DL (FinBERT)	Unstructured	FinBERT	F1: 0.81-0.87	10.6M notes	Finnish-only model
[52]More (2024)	Medication extraction	Rule-based	Unstructured	SpaCy + Med7	Example outputs	Simulated	No metrics, proof of concept
[38]Alskaf et al. (2024)	Cardiac outcome prediction	Rule-based + DL	Unstructured + CMR	MedCAT, SVM	AUC: 0.82	4,188 patients	Manual CMR data
[53]Badalotti et al. (2024)	Comorbidity extraction	AL + BERT	Unstructured	BERT + BALD	F1 > 0.70	561 stroke patients	No negation, rare condition limits
[54]Martina et al. (2024)	SDOH (employment, language)	Rule-based	Unstructured	Regex	PPV: 88-99%	17,838 admissions	No sensitivity calc
[55]Ruckdeschel et al. (2025)	Smoking history + LDCT	Rule-based + DL	Unstructured	BlueBERT + Rules	F1: 0.88	MIMIC-III	No external validation
[56]Clay et al. (2025)	NLP methods overview	Statistical + DL	Unstructured	BERT, GPT	Up to F1: 0.92	Multiple case studies	LLM explainability, hallucination
[57]Chen et al. (2025)	NLP bibliometric review	CiteSpace, VOSviewer	Scientific Literature	———	Trend analysis only	4,803 papers	No performance metrics
[58]Takeuchi et al. (2025)	Tumor response prediction	DL (LSTM, MTL)	Unstructured	LSTM	F1: 0.72	716 treatments	No external validation
[59]Bryde Alnor et al. (2025)	Bleeding risk prediction	Regex NLP	Unstructured + Labs	TensorFlow	C-stat: 0.72	46,439 patients	Single center, missing data
[60]Soto Jacome et al. (2024)	TUS appropriateness audit	Rule-based	Unstructured	sNLP, aNLP	F1: 0.97-0.99	628 patients	Data imbalance
[24]Li et al. (2025)	Suicide risk prediction	AMC + NLP	Unstructured	AMC + SEANCE	F1: 0.20	25,342 VA pts	VA only, high compute
[39]Pan et al. (2025)	Multi-condition detection	LLM + Rules	Unstructured	Mistral-7B	Sens: 88-94%	551K notes	Cardiac-only cohort

4. Synthesis Of NLP Applications in Healthcare

Recent advancements in natural language processing (NLP) have instigated a transformative shift in healthcare research, evidenced by various studies employing rule-based systems, statistical learning, deep learning (DL), and large language models (LLMs) to address numerous clinical challenges. An incisive analysis of these initiatives uncovers both encouraging advancements and persistent constraints that influence the trajectory of forthcoming research.

1. Varied Tasks and Methodological Selections

The range of NLP applications is extensive, encompassing smoking status classification [41] ,suicide risk prediction [44]; [24], and extraction of Social Determinants of Health [48]; [46], as well as specialized tasks such as Dravet syndrome detection [43] and tumor response prediction [23].

A distinct methodological pattern is evident: simpler, rule-based techniques prevail in contexts necessitating high interpretability or where domain-specific terminology is comprehensively understood [33];[42], whereas deep learning and transformer-based models are favored for more intricate or extensive information extraction[61].Despite the growing trend towards deep learning, particularly BERT variants and CNN/LSTM hybrids, numerous studies persist in utilizing rule-based pipelines or integrating them with machine learning to enhance precision and transparency, especially in clinically sensitive contexts [38]; [55].

Nevertheless, although these hybrid methods yield practical accuracy, they frequently lack generalizability, a limitation intensified by the dominance of single-site data sources.

2. Data Type and Scale: Advantages and Disadvantages

A significant advantage in numerous studies is the availability of real-world electronic health records (EHRs), encompassing structured, semi-structured, and unstructured data. However, this also presents significant challenges. Numerous high-performing models [48], [51]rely on extensive datasets, which, although advantageous for training, pose issues regarding reproducibility and accessibility, many datasets are still confined or proprietary.

Furthermore, research utilizing limited data sets [41][43] frequently indicates exaggerated performance in internal validations while experiencing challenges in generalization. This outlook indicates an excessive dependence on in-domain tuning, accompanied by minimal attempts at external validation or cross-institutional testing. Despite extensive datasets [39],[49] assumptions regarding pristine input or restricted clinical cohorts constrain wider applicability.

3. Performance Metrics and Model Validation

The F1-score and AUC are the predominant metrics in various studies, particularly in classification tasks. Elevated reported scores [44], with AUC 0.93;[60] with F1 up to 0.99) are promising but require careful interpretation. These metrics frequently originate from internally compiled datasets, and in the absence of external validation, their practical applicability is dubious.Certain studies[56]recognize the constraints of large language models (LLMs) such as GPT in clinical environments, particularly regarding hallucination and insufficient explainability. This critical reflection is a significant contribution, emphasizing that performance metrics are inadequate without comprehending model behavior, particularly in care decisions.

4. Constraints and Prospects for Progress

Several persistent constraints arise:

Dataset limitations: Numerous models are developed using small, localized datasets, which restrict their generalizability [47],[50]

Insufficient external validation: Most models are site-specific or developed within singular health systems [55][59], limiting their wider applicability.

Clinical scope: Numerous studies concentrate exclusively on populations or conditions, such as veterans [24] or cardiac patients[39], hindering broader applicability.

Negation and context management: Certain approaches, especially rule-based ones, encounter difficulties with clinical subtleties, including negation identification and ambiguity in temporal references [53].Nevertheless, the domain is not devoid of advancement. Research by [48] and[58]illustrates that multi-modal data fusion and multi-task learning enhance clinical outcome prediction and entity recognition. Applying active learning (AL) and domain-specific transformers, such as mSpERT and FinBERT, demonstrates promise for effective, scalable, and contextually aware natural language processing (NLP) solutions.

5. EXTRACTED STATISTICS

Figure (1) depicts the allocation of NLP methods employed in healthcare research from 2020 to 2025, indicating that deep learning (DL) approaches are the most dominant, comprising roughly 38% of the studies. This dominance signifies the increasing dependence on sophisticated models such as BERT, LSTM, and CNN for intricate tasks like information extraction and risk prediction. Rule-based methods, though conventional, constitute approximately 29% of the total, highlighting their sustained significance owing to their interpretability and efficacy in clearly defined clinical scenarios. Hybrid methodologies that amalgamate rule-based logic with deep learning yield approximately 14%, indicating a trend towards integrating precision and scalability. Alternative methodologies, such as regular expressions, active learning, bibliometric instruments, and hybrids based on large language models (LLMs), constitute smaller yet burgeoning segments. The chart underscores a developing domain where the necessity for explainability and practical utility in real-world healthcare settings balances advancements in model complexity.

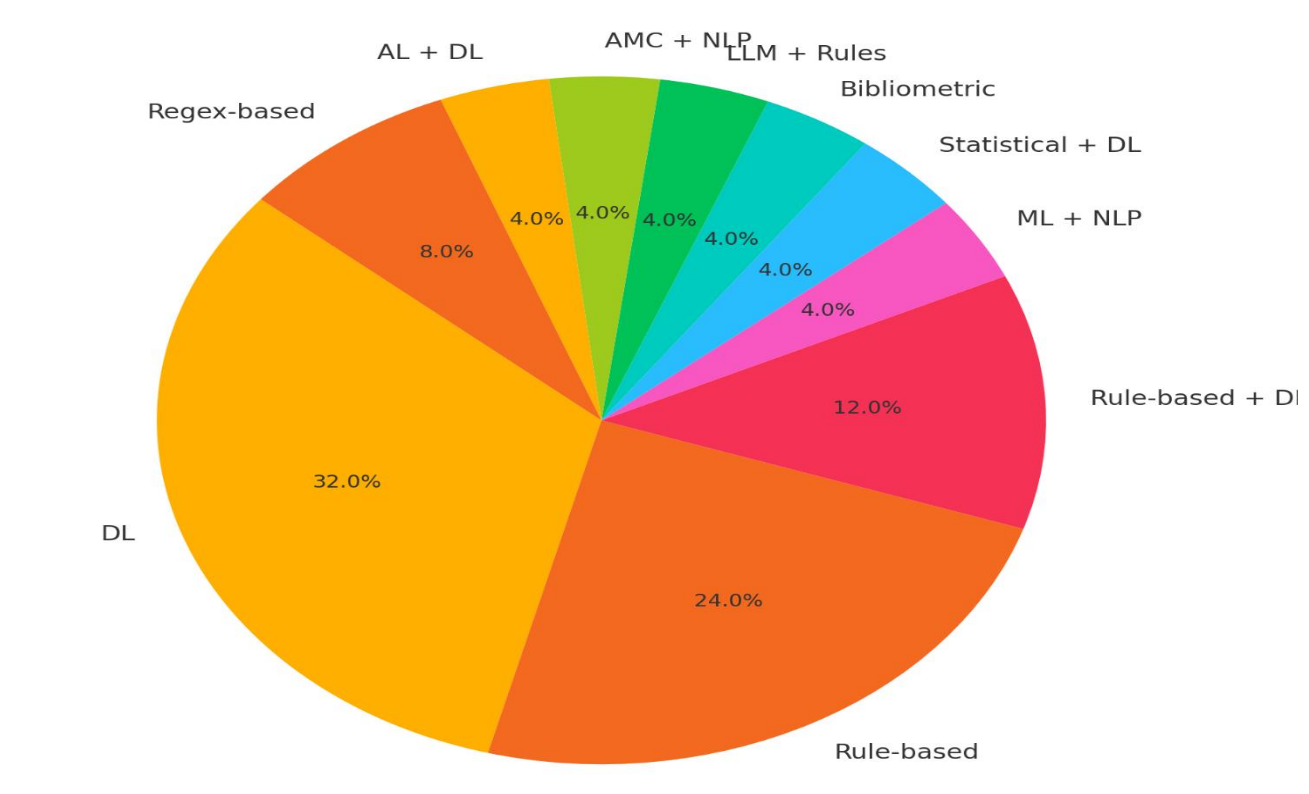

Figure 1. Statistical representation of NLP papers (2020 – 2025) based on the NLP methods

Figure (2) represents the allocation of various NLP model types utilised in a series of studies. The predominant methodologies employed are BERT and its derivatives (including FinBERT, BlueBERT, and mSpERT) and Rule-based Systems, each representing 22.2% of the total. This standpoint signifies an equilibrium between conventional rule-based methodologies and contemporary transformer-based models. CNN, LSTM, and other deep learning models account for 18.5%, indicating a persistent dependence on deep learning architectures. Hybrid models, integrating rules and learning techniques (e.g., DL, MTL, AMC, AL), along with alternative methods (such as SpaCy, Watson, and custom Java pipelines), each account for 14.8%, indicating various integrative strategies. Ultimately, LLMs (e.g., GPT, Mistral-7B) constitute merely 7.4% of usage, indicating that although large language models are increasing in popularity, they have not yet achieved dominance in these specific studies.

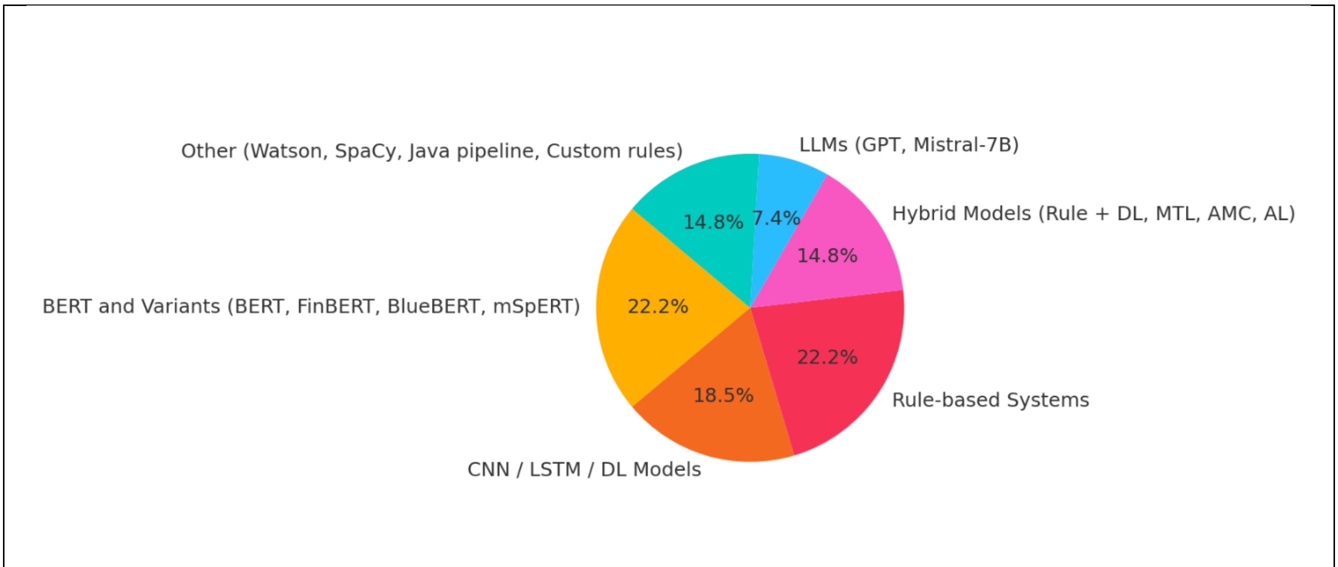

Figure 2. Statistical representation of NLP papers (2020 – 2025) based on the Model

Figure (3) indicates that a significant proportion of NLP research in healthcare from 2020 to 2025 attained high performance, with around 41% reporting F1-scores of 0.85 or greater or AUCs exceeding 0.90. This perspective demonstrates the efficacy of advanced models, especially deep learning methodologies, in managing intricate clinical text tasks. Moderate performance models, accounting for approximately 27%, indicate the necessity for continued efforts to enhance accuracy in tasks hindered by data constraints or domain-specific intricacies. Simultaneously, underperforming models (approximately 9%) underscore the constraints of specific methodologies or resource-limited contexts. Significantly, approximately 23% of studies failed to present quantitative metrics, depending instead on qualitative assessments or descriptive results, whereas 9% were exploratory or trend-oriented without formal performance validation. The chart highlights advancements in developing high-performing clinical NLP systems while emphasizing the necessity for more uniform and stringent evaluation standards within the domain.

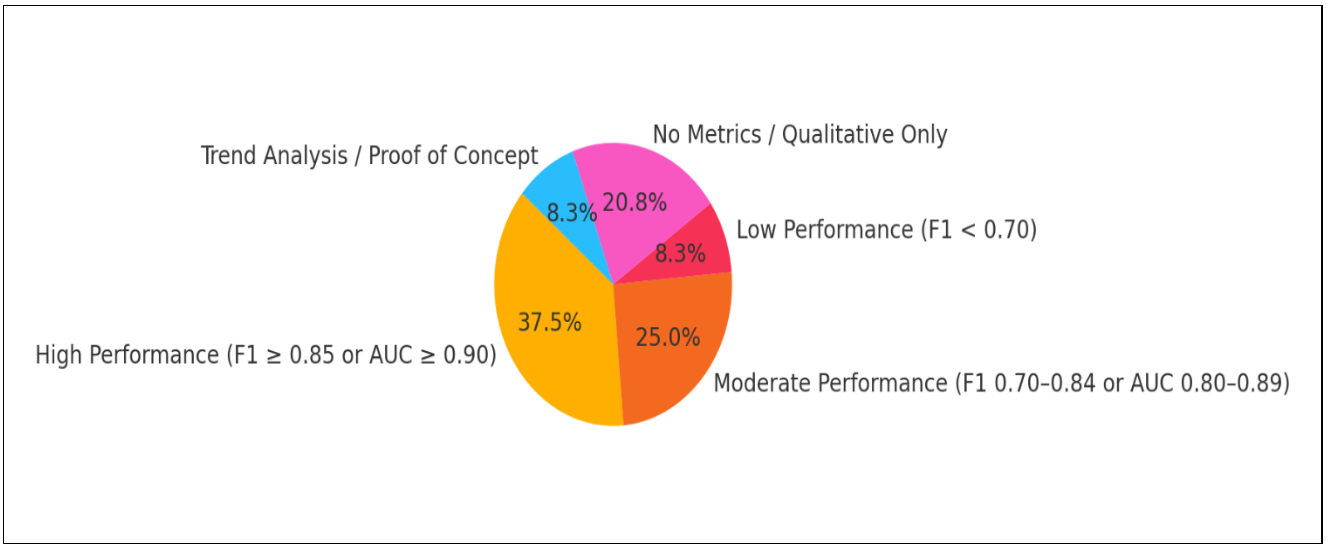

Figure 3. Statistical representation of NLP papers (2020 – 2025) based on the Performance

6. RECOMMENDATIONS & RESEARCH AGENDA

Given the present condition of NLP applications in healthcare, various strategic recommendations and research avenues arise to improve impact, equity, and clinical efficacy. The recommendations are structured around five principal dimensions: generalizability, validation, interpretability, data diversity, and integration into clinical workflows.

1. Expand Dataset Heterogeneity and External Verification

Recommendation: Future research should extend beyond single-institution or condition-specific datasets to encompass multi-site, demographically diverse populations.

Rationale: Numerous studies [44],[41],[47]utilize restricted samples, constraining generalizability. External validation on novel, diverse datasets should be established as a standard for model publication and deployment readiness.

Research Agenda: Develop or enhance federated or open-access clinical NLP datasets with annotations representing varied patient demographics.

Establish resilient domain adaptation methodologies to refine models across institutions without necessitating complete retraining.

Examine zero-shot or few-shot learning in the context of rare disease documentation [50].

2. Incorporate Interpretability and Clinical Transparency

Recommendation: Priorities models that facilitate clinical interpretability or incorporate human-in-the-loop frameworks to ensure accountability in decision-making.

Rationale: As [56] emphasize, the opacity of LLMs (e.g., GPT, BERT) presents a considerable obstacle to clinical trust. Clinicians must comprehend the reasoning underlying model outputs to make informed decisions.

Research Agenda: Create interpretable NLP models that emphasise significant features, sentence segments, or reasoning pathways.

Investigate the application of contrastive explanations or attention visualization to elucidate entity recognition or classification decisions.

Assess clinician trust and adoption in models both with and without explainability features.

3. Adopt Multimodal and Multitask Learning

Recommendation: Utilise multimodal data (e.g., imaging, laboratory

results, unstructured text) and multitask architectures to encapsulate the intricacies of clinical decision-making.

Rationale: Research by [38] and [58] demonstrates that integrating supplementary data sources beyond textual information can enhance predictive accuracy, particularly in outcome modelling.

Research Agenda: Develop and evaluate integrated models that acquire knowledge from text, imagery, structured laboratory data, and temporal patterns.

Investigate multitask learning in natural language processing to concurrently extract entities, execute classification, and acquire temporal information[62].

Evaluate the practical implications of these models on diagnostic or triage workflows.

4. Tackle Bias, Negation, and Linguistic Complexity

Suggestion: Enhance the resilience of NLP to the intricacies of clinical language, encompassing negation, uncertainty, and social/cultural bias.

Justification: Rule-based systems and deep learning models frequently struggle with nuanced or negated statements [53][43], and demographic biases can distort prediction results.

Research Agenda: Create negation-aware transformers or modify models such as NegBio to integrate into contemporary pipelines[61].

Assess models for equity concerning age, gender, language proficiency, and race in extracting social determinants of health[46];[54].

Evaluate NLP tools on marginalized languages and multilingual clinical documentation (e.g., the application of FinBERT for Finnish by[51].

5. Advocate for Practical Usability and Human-Centric Assessment

Recommendation: Anchor NLP development in clinical applications by engaging healthcare professionals during the design and testing phases.

Rationale: Numerous studies indicate robust in-sample performance (e.g., F1 > 0.85), yet few evaluate model efficacy in real-world applications. Adoption remains constrained unless the impact on clinician workload, diagnostic quality, and patient safety is assessed.

Research Agenda: Execute prospective studies or A/B testing of NLP tools in active clinical environments.

Create user-centric design frameworks for integrating natural language processing (NLP) into electronic health records (EHRs).

Examine cognitive load, clinician satisfaction, and task alignment during implementation of NLP tools at the point of care.

7. CONCLUSIONS

Natural language processing (NLP) has become a critical tool to take advantage of the huge volumes of structured and unstructured data in Electronic Health Records (EHRs). This review covered a vast number of studies illustrating the ways in which NLP facilitates automated extraction of information, disease identification, and decision support in the clinical domain. From the pioneering rule-based systems to sophisticated deep learning and transformers, the research has made tremendous advances in approach and application. Techniques such as BERT, CNNs, and LSTMs have been strongly performing in extracting sophisticated medical concepts from clinical narratives. Such methodologies have been found useful in applications such as classifying smoking status, social determinants of health identification, and preventive screening eligible patients.

Despite all these advances, there remain several challenges to widespread adoption. Clinical texts remain very heterogeneous, with vocabulary and terminology that vary between institutions. The scarcity of annotated datasets also impacts training the models, particularly for uncommon diseases and underserved populations. Furthermore, the lack of explainability of deep learning models presents challenges in the clinical context, in which explainability is paramount. Most of the existing NLP systems lack generalizability across varied healthcare settings. Integration into real-time clinical practices continues to be restricted by the technical, ethical, and regulatory barriers.

In the future, research should be directed towards developing explainable, yet scalable and responsible, NLP architectures. Enhancing collaboration between clinicians, data scientists, and informaticians will be the lynchpin to facilitating real-world implementation. With ongoing innovation, NLP could revolutionize data-driven medicine and achieve more personalized, efficient, and equitable care delivery.

REFERENCES

1. D. H. Maulud *et al.*, "Review on Natural Language Processing Based on Different Techniques," *Asian Journal of Research in Computer Science*, pp. 1–17, Jun. 2021, doi: 10.9734/ajrcos/2021/v10i130231.
2. R. M. Abdullah *et al.*, "Paralinguistic Speech Processing: An Overview," *Asian Journal of Research in Computer Science*, pp. 34–46, Jun. 2021, doi: 10.9734/ajrcos/2021/v10i130233.
3. N. R *et al.*, "Enhancing drug discovery and patient care through advanced analytics with the power of NLP and machine learning in pharmaceutical data interpretation," *SLAS Technol*, vol. 31, p. 100238, Apr. 2025, doi: 10.1016/J.SLAST.2024.100238.
4. V. A. Batista and A. G. Evsukoff, "Application of Transformers based methods in Electronic Medical Records: A Systematic Literature Review," Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.02768
5. C. H. Chang *et al.*, "New opportunities and challenges for conservation evidence synthesis from advances in natural language processing," *Conservation Biology*, vol. 39, no. 2, Apr. 2025, doi: 10.1111/cobi.14464.

6. J. Ye, L. He, J. Hai, C. Xu, S. Ding, and M. Beestrum, "Development and Application of Natural Language Processing on Unstructured Data in Hypertension: A Scoping Review," Feb. 29, 2024. doi: 10.1101/2024.02.27.24303468.
7. E. Hossain *et al.*, "Natural Language Processing in Electronic Health Records in Relation to Healthcare Decision-making: A Systematic Review," Jun. 2023, [Online]. Available: http://arxiv.org/abs/2306.12834
8. M. Omar, S. Nassar, K. SharIf, B. S. Glicksberg, G. N. Nadkarni, and E. Klang, "Emerging applications of NLP and large language models in gastroenterology and hepatology: a systematic review," *Front Med (Lausanne)*, vol. 11, Jan. 2025, doi: 10.3389/fmed.2024.1512824.
9. L. Lilli *et al.*, "A Comprehensive Natural Language Processing Pipeline for the Chronic Lupus Disease," *Stud Health Technol Inform*, vol. 316, pp. 909–913, Aug. 2024, doi: 10.3233/SHTI240559.
10. R. Gatsinga *et al.*, "Current Applications and Developments of Natural Language Processing in Kidney Transplantation: A Scoping Review," *Transplant Proc*, vol. 57, no. 4, pp. 558–568, May 2025, doi: 10.1016/J.TRANSPROCEED.2025.02.027.
11. J. Song, S. Oh, L. Evans, K. H. Bowles, and A. Chair, "Social Risk Factors are Associated with Risk for Hospitalization in Home Health Care: A Natural Language Processing Study," 2023.
12. T. A. Koleck, C. Dreisbach, P. E. Bourne, and S. Bakken, "Natural language processing of symptoms documented in free-text narratives of electronic health records: A systematic review," Feb. 06, 2019, *Oxford University Press*. doi: 10.1093/jamia/ocy173.
13. A. Kormilitzin, N. Vaci, Q. Liu, and A. Nevado-Holgado, "Med7: a transferable clinical natural language processing model for electronic health records," Mar. 2020, [Online]. Available: http://arxiv.org/abs/2003.01271
14. I. Li *et al.*, "Neural Natural Language Processing for Unstructured Data in Electronic Health Records: a Review," Jul. 2021, [Online]. Available: http://arxiv.org/abs/2107.02975
15. M. Nancy, A. M. Nancy, and R. Maheswari, "A REVIEW ON UNSTRUCTURED DATA IN MEDICAL DATA," 200AD. [Online]. Available: https://www.researchgate.net/publication/343655656
16. A. Guazzo, E. Longato, G. P. Fadini, M. L. Morieri, G. Sparacino, and B. Di Camillo, "Deep-learning-based natural-language-processing models to identify cardiovascular disease hospitalisations of patients with diabetes from routine visits' text," *Sci Rep*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-45115-1.
17. L. M. Peltonen, H. von Gerich, E. Myllymäki, J. Walsh, and M. Medvecky, "Exploring Delays in Cardiac Care Processes Through Electronic Health Records," *Stud Health Technol Inform*, vol. 316, pp. 1866–1870, Aug. 2024, doi: 10.3233/SHTI240795.
18. N. Cook *et al.*, "Engaging stakeholders with professional or lived experience to improve firearm violence NLP lexicon (Preprint)," Oct. 29, 2024. doi: 10.2196/preprints.68105.
19. M. Russel Hossain, S. Mahabub, A. Al Masum, and I. Jahan, "Natural Language Processing (NLP) in Analyzing Electronic Health Records for Better Decision Making," 2024, doi: 10.32996/jcsts.
20. R. Shankar, A. Bundele, and A. Mukhopadhyay, "Natural language processing of electronic health records for early detection of cognitive decline: a systematic review," *NPJ Digit Med*, vol. 8, no. 1, Dec. 2025, doi: 10.1038/s41746-025-01527-z.
21. J. Shi *et al.*, "Identifying Patients Who Meet Criteria for Genetic Testing of Hereditary Cancers Based on Structured and Unstructured Family Health History Data in the Electronic Health Record: Natural Language Processing Approach," *JMIR Med Inform*, vol. 10, no. 8, Aug. 2022, doi: 10.2196/37842.
22. A. Houston, S. Williams, W. Ricketts, C. Gutteridge, C. Tackaberry, and J. Conibear, "Automated derivation of diagnostic criteria for lung cancer using natural language processing on electronic health records: a pilot study," *BMC Medical Informatics and Decision Making 2024 24:1*, vol. 24, no. 1, pp. 1–10, Dec. 2024, doi: 10.1186/S12911-024-02790-Y.
23. P. Thatoi, A. Shiwlani, H. Qureshi, R. Choudhary, H. A. Qureshi, and S. Kumar, "Natural Language Processing (NLP) in the Extraction of Clinical Information from Electronic Health Records (EHRs) for Cancer Prognosis," 2023. [Online]. Available: https://www.researchgate.net/publication/383650951
24. S. Li *et al.*, "Preprocessing of natural language process variables using a data-driven method improves the association with suicide risk in a large veterans affairs population," *Comput Biol Med*, vol. 189, May 2025, doi: 10.1016/j.compbiomed.2025.109939.
25. G. A. Ismael *et al.*, "Scheduling Algorithms Implementation for Real Time Operating Systems: A Review," *Asian Journal of Research in Computer Science*, pp. 35–51, Sep. 2021, doi: 10.9734/ajrcos/2021/v11i430269.
26. S. T. Ibrahim, M. Li, J. Patel, and T. R. Katapally, "Utilizing natural language processing for precision prevention of mental health disorders among youth: A systematic review," *Comput Biol Med*, vol. 188, Apr. 2025, doi: 10.1016/j.compbiomed.2025.109859.

27. I. Mahmood Ibrahim *et al.*, "Task Scheduling Algorithms in Cloud Computing: A Review," 2021.
28. D. M. Abdullah *et al.*, "Secure Data Transfer over Internet Using Image Steganography: Review," *Asian Journal of Research in Computer Science*, pp. 33–52, Jul. 2021, doi: 10.9734/ajrcos/2021/v10i330243.
29. Y. Wieland-Jorna, D. Van Kooten, R. A. Verheij, Y. De Man, A. L. Francke, and M. G. Oosterveld-Vlug, "Natural language processing systems for extracting information from electronic health records about activities of daily living. A systematic review," Jul. 01, 2024, *Oxford University Press*. doi: 10.1093/jamiaopen/ooae044.
30. J. Zeng *et al.*, "Natural Language Processing to Identify Cancer Treatments With Electronic Medical Records," 2021. [Online]. Available: https://doi.org/10.
31. Z. Zeng, Y. Deng, X. Li, T. Naumann, and Y. Luo, "Natural Language Processing for EHR-Based Computational Phenotyping," *IEEE/ACM Trans Comput Biol Bioinform*, vol. 16, no. 1, pp. 139–153, Jan. 2019, doi: 10.1109/TCBB.2018.2849968.
32. M. K. S. Uddin, "A REVIEW OF UTILIZING NATURAL LANGUAGE PROCESSING AND AI FOR ADVANCED DATA VISUALIZATION IN REAL-TIME ANALYTICS," *Global Mainstream Journal*, vol. 1, no. 4, pp. 34–49, Jul. 2024, doi: 10.62304/ijmisds.v1i04.185.
33. A. S. Lacey *et al.*, "Obtaining structured clinical data from unstructured data using natural language processing software," *Int J Popul Data Sci*, vol. 1, no. 1, Apr. 2017, doi: 10.23889/ijpds.v1i1.381.
34. K. Du, Y. Zhao, R. Mao, F. Xing, and E. Cambria, "Natural language processing in finance: A survey," Mar. 01, 2025, *Elsevier B.V.* doi: 10.1016/j.inffus.2024.102755.
35. A. Zubiaga, "Natural language processing in the era of large language models," *Front Artif Intell*, vol. 6, 2023, doi: 10.3389/frai.2023.1350306.
36. D. H. Maulud *et al.*, "Review on Natural Language Processing Based on Different Techniques," *Asian Journal of Research in Computer Science*, pp. 1–17, Jun. 2021, doi: 10.9734/ajrcos/2021/v10i130231.
37. S. Golder, D. Xu, K. O'Connor, Y. Wang, M. Batra, and G. G. Hernandez, "Leveraging Natural Language Processing and Machine Learning Methods for Adverse Drug Event Detection in Electronic Health/Medical Records: A Scoping Review," Apr. 01, 2025, *Adis*. doi: 10.1007/s40264-024-01505-6.
38. E. Alskaf *et al.*, "Machine learning outcome prediction using stress perfusion cardiac magnetic resonance reports and natural language processing of electronic health records," *Inform Med Unlocked*, vol. 44, Jan. 2024, doi: 10.1016/j.imu.2023.101418.
39. J. Pan *et al.*, "Integrating large language models with human expertise for disease detection in electronic health records," *Comput Biol Med*, vol. 191, Jun. 2025, doi: 10.1016/j.compbiomed.2025.110161.
40. A. Kormilitzin, N. Vaci, Q. Liu, and A. Nevado-Holgado, "Med7: a transferable clinical natural language processing model for electronic health records," Mar. 2020, [Online]. Available: http://arxiv.org/abs/2003.01271
41. S. Rajendran and U. Topaloglu, "Extracting Smoking Status from Electronic Health Records Using NLP and Deep Learning."
42. M. D. Solomon, G. Tabada, A. Allen, S. H. Sung, and A. S. Go, "Large-scale identification of aortic stenosis and its severity using natural language processing on electronic health records," *Cardiovasc Digit Health J*, vol. 2, no. 3, pp. 156–163, Jun. 2021, doi: 10.1016/j.cvdhj.2021.03.003.
43. T. Lo Barco, M. Kuchenbuch, N. Garcelon, A. Neuraz, and R. Nabbout, "Improving early diagnosis of rare diseases using Natural Language Processing in unstructured medical records: an illustration from Dravet syndrome," *Orphanet J Rare Dis*, vol. 16, no. 1, Dec. 2021, doi: 10.1186/s13023-021-01936-9.
44. F. R. Tsui *et al.*, "Natural language processing and machine learning of electronic health records for prediction of first-time suicide attempts," *JAMIA Open*, vol. 4, no. 1, Jan. 2021, doi: 10.1093/jamiaopen/ooab011.
45. E. Hatef *et al.*, "Development and assessment of a natural language processing model to identify residential instability in electronic health records' unstructured data: A comparison of 3 integrated healthcare delivery systems," *JAMIA Open*, vol. 5, no. 1, Apr. 2022, doi: 10.1093/jamiaopen/ooac006.
46. S. Han *et al.*, "Classifying social determinants of health from unstructured electronic health records using deep learning-based natural language processing," Mar. 01, 2022, *Academic Press Inc.* doi: 10.1016/j.jbi.2021.103984.
47. D. H. P. Benício, J. C. Xavier Junior, K. R. S. de Paiva, and J. D. de A. S. Camargo, "Applying Text Mining and Natural Language Processing to Electronic Medical Records for extracting and transforming texts into structured data," *Research, Society and Development*, vol. 11, no. 6, p. e37711629184, Apr. 2022, doi: 10.33448/rsd-v11i6.29184.
48. K. Lybarger *et al.*, "Leveraging natural language processing to augment structured social determinants of health data in the electronic health record," *J Am*

Med Inform Assoc, vol. 30, no. 8, pp. 1389–1397, Aug. 2023, doi: 10.1093/jamia/ocad073.
49. S. Liu *et al.*, "Leveraging natural language processing to identify eligible lung cancer screening patients with the electronic health record," *Int J Med Inform*, vol. 177, Sep. 2023, doi: 10.1016/j.ijmedinf.2023.105136.
50. H. Barman *et al.*, "Retrospective study of propionic acidemia using natural language processing in Mayo Clinic electronic health record data," *Mol Genet Metab*, vol. 140, no. 3, Nov. 2023, doi: 10.1016/j.ymgme.2023.107695.
51. J. Lin *et al.*, "Identification of Health Conditions in Unstructured Health Records with Deep Learning-Based Natural Language Processing," Oct. 10, 2024. doi: 10.1101/2024.10.08.24315141.
52. A. A. More, "Natural Language Processing - Based Structured Data Extraction from Unstructured Clinical Notes," *Journal of Contemporary Medical Practice*, vol. 6, no. 8, pp. 327–330, Aug. 2024, doi: 10.53469/jcmp.2024.06(08).67.
53. D. Badalotti, A. Agrawal, U. Pensato, G. Angelotti, and S. Marcheselli, "Development of a Natural Language Processing (NLP) model to automatically extract clinical data from electronic health records: results from an Italian comprehensive stroke center," *Int J Med Inform*, vol. 192, Dec. 2024, doi: 10.1016/j.ijmedinf.2024.105626.
54. E. A. Martin, A. G. D'Souza, V. Saini, K. Tang, H. Quan, and C. A. Eastwood, "Extracting social determinants of health from inpatient electronic medical records using natural language processing," *Journal of epidemiology and population health*, vol. 72, no. 6, Dec. 2024, doi: 10.1016/j.jeph.2024.202791.
55. J. C. Ruckdeschel *et al.*, "Unstructured Data Are Superior to Structured Data for Eliciting Quantitative Smoking History From the Electronic Health Record," 2023, doi: 10.1200/CCI.22.
56. B. Clay, H. I. Bergman, S. Salim, G. Pergola, J. Shalhoub, and A. H. Davies, "Natural language processing techniques applied to the electronic health record in clinical research and practice - an introduction to methodologies," Apr. 01, 2025, *Elsevier Ltd*. doi: 10.1016/j.compbiomed.2025.109808.
57. X. Chen, W. Tian, and H. Fang, "Bibliometric analysis of natural language processing using CiteSpace and VOSviewer," *Natural Language Processing Journal*, vol. 10, p. 100123, Mar. 2025, doi: 10.1016/j.nlp.2024.100123.
58. T. Takeuchi *et al.*, "A series of natural language processing for predicting tumor response evaluation and survival curve from electronic health records," *BMC Med Inform Decis Mak*, vol. 25, no. 1, p. 85, Dec. 2025, doi: 10.1186/s12911-025-02928-6.
59. A. B. Alnor, R. B. Lynggaard, M. S. Laursen, and P. J. Vinholt, "Natural language processing for identifying major bleeding risk in hospitalised medical patients," *Comput Biol Med*, vol. 190, May 2025, doi: 10.1016/j.compbiomed.2025.110093.
60. C. Soto Jacome *et al.*, "Thyroid Ultrasound Appropriateness Identification Through Natural Language Processing of Electronic Health Records," *Mayo Clinic Proceedings: Digital Health*, vol. 2, no. 1, pp. 67–74, Mar. 2024, doi: 10.1016/j.mcpdig.2024.01.001.
61. B. G. Patra *et al.*, "Extracting social determinants of health from electronic health records using natural language processing: A systematic review," Dec. 01, 2021, *Oxford University Press*. doi: 10.1093/jamia/ocab170.
62. M. Russel Hossain, S. Mahabub, A. Al Masum, and I. Jahan, "Natural Language Processing (NLP) in Analyzing Electronic Health Records for Better Decision Making," 2024, doi: 10.32996/jcsts.