Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 07 Aug 2026
Revised 11 Sep 2026
Accepted 29 Sep 2026

Natural Resources for Human Health 2026, Vol. 6 (23s): 587–605

KEYWORDS:
Natural Language Processing; Electronic Health Records; Clinical Notes; Cancer Research; Information Extraction; Text Classification; Machine Learning; Deep Learning; Clinical Text Mining; Chronic Mobility Disability; Healthcare Analytics.

NLP-based Intelligent Extraction and Analysis of Electronic Health Records for Cancer Research and Chronic Mobility Disability Identification

Jagrati Shekhawat[1], Dr. Viral Parekh[2]

[1]Research Scholar, Monark University, Computer Engineering Gujrat, India
[2]Associate Professor, Department of Computer Engineering, Monark University.

ABSTRACT: Electronic Health Records (EHRs) contain a large amount of patient information. Some of this information is stored in structured fields such as diagnosis codes, laboratory values, medications, procedures, and demographic details. A large amount is also written as free text in physician notes, discharge summaries, pathology reports, radiology reports, and other clinical documents. This free text often contains useful information that is difficult and time-consuming to extract manually. This study examines AI-enhanced Natural Language Processing (NLP) for EHR analysis from two connected perspectives. First, it presents a systematic review of NLP applications in cancer research. Second, it examines an EHR-based example of NLP screening for chronic mobility disability among colorectal cancer patients[1]. The literature review used Scopus and included peer-reviewed English-language journal studies published from 2019 to 2024. The search identified 272 records; after screening and eligibility assessment, 94 studies were included.

The reviewed studies commonly addressed breast, lung, and colorectal cancer. Information extraction and text classification were frequent NLP tasks, while the literature showed a gradual movement from rule-based and keyword methods toward machine learning, deep learning, and transformer-based approaches. The mobility example included 14,877 colorectal cancer patients and approximately 303,182 clinical notes. Keyword screening identified 1,482 notes from 420 patients. After contextual manual review, 286 notes involving 105 patients contained information about both the reason and duration of wheelchair use.

The findings show that NLP can reduce the amount of clinical text that researchers need to review manually, but keyword matching alone cannot reliably confirm chronic mobility disability. Context, negation, time, patient reference, and clinical meaning must be considered. This study therefore presents a practical framework for future context-aware, explainable, privacy-preserving, and externally validated clinical NLP systems. The work is an evidence synthesis and methodological framework; it does not claim a newly trained predictive model or new model-performance scores.

1. INTRODUCTION

1.1 Electronic Health Records

EHRs have become a major source of information for healthcare research. They record medical history, diagnoses, symptoms, laboratory results, medications, procedures, treatments, and outcomes. EHR information can be broadly viewed as structured and unstructured data. Structured information follows predefined formats and is usually easier to search[2]. Examples include ICD diagnosis codes, laboratory values, medication records, procedure codes, and demographic fields.

Unstructured clinical text is different. It is written naturally by healthcare professionals and may contain detailed observations that are not captured in coded fields. Physician notes, nursing notes, discharge summaries, consultation notes, pathology reports, and radiology reports can describe symptoms, treatment response, disease progression, functional limitations, and other details. For research, these narrative records can be especially valuable because they provide clinical context[2].

1.2 Why Clinical Text Is Difficult

Clinical language is not written like ordinary text. Clinicians often use abbreviations, short phrases, technical terms, incomplete sentences, and local documentation styles. The same condition may be described in several ways. For example, mobility problems may appear as difficulty walking, impaired gait, limited mobility, inability to walk independently, wheelchair dependence, or need for assistance. These expressions are related, but they may not have exactly the same meaning[3].

A word can also have different meanings depending on context. A note may say that a patient uses a wheelchair, that a wheelchair is planned after surgery, that a family member uses a wheelchair, or that a wheelchair was used temporarily. A simple keyword search treats these situations similarly even though their clinical meaning is very different. This is a central challenge for clinical NLP.

1.3 Role of NLP

NLP is a branch of Artificial Intelligence that helps computers process human language. In healthcare, clinical NLP can convert free text into information that can be searched, compared, classified, and studied. It can reduce the number of records that require detailed manual review and can support the creation of research cohorts.

Common clinical NLP tasks include named entity recognition, information extraction, relation extraction, concept normalization, text classification, temporal information extraction, and de-identification. The choice of technique depends on the research problem, available data, and required level of accuracy.

1.4 Study Focus

This paper focuses on two related uses of NLP: cancer research and chronic mobility disability identification. Both require meaningful information to be extracted from large collections of clinical text. The first part provides evidence from published research[3]. The second part illustrates how EHR notes can be screened for mobility-related information and then manually validated.

2. BACKGROUND AND RELATED WORK

2.1 NLP in Cancer Research

Cancer-related information is often spread across several documents. A pathology report may contain diagnosis information, an oncology note may describe stage and treatment decisions, a radiology report may describe imaging findings, and a follow-up note may discuss treatment response or recurrence. Structured EHR fields alone may not contain all of these details[3]. NLP can connect information across documents and convert narrative information into research-ready variables.

The reviewed literature shows several important applications. NLP can help identify patients with specific cancer characteristics, extract symptoms and clinical findings, identify cancer stage, capture treatments and medications, study treatment response, detect recurrence, and support prognosis and outcome research. It can also assist with clinical trial eligibility and database construction.

2.2 Examples Across Cancer Types

Cancer area	Examples of NLP use	Reported example
Breast	Trial eligibility, database construction, treatment and palliative-care information	Used to identify and organise relevant clinical information
Prostate	Metastatic disease, TNM stage, examinations, treatment guidance, urinary and social outcomes	Supports extraction of clinical and patient-related information
Ovarian	Clinical audit and post-surgery outcome	RoBERTa example reported about 81%

	analysis	accuracy
Cervical	Cancer-stage classification from MRI, examination and PET-CT reports	Example reported about 94% accuracy
Endocrine/thyroid	Diagnosis, stage and thyroid-nodule characteristics	GatorTron examples reported F1 values of 0.885–0.950
Skin	Pathology information for skin cancer and melanoma	Melanoma example reported a 90.4% matching rate
CNS	MRI-report analysis and brain-metastasis identification	Example reported about 83% accuracy and AUROC 0.92

These reported results should not be treated as directly comparable because the studies used different datasets, populations, cancer types, annotation procedures, tasks, and evaluation measures.

2.3 Development of NLP Methods

The literature suggests a broad technology progression. Earlier systems often used keyword lists, dictionaries, and manually written rules. These approaches are simple and can be useful when the target terms are stable. Traditional machine learning then allowed models to learn patterns from labelled examples. Deep learning extended this capability, and transformer-based clinical language models have more recently become important because they can represent relationships between words and sentences[4-5].

This progression does not mean that newer models are always better for every task. A simple rule-based approach may be appropriate for an initial screening stage, while a context-aware model may be needed for ambiguous clinical language. The best approach depends on the purpose, data quality, annotation availability, and required reliability.

3. RESEARCH OBJECTIVES AND RESEARCH QUESTIONS

3.1 Objectives

- Examine how NLP is used to analyse EHRs and clinical notes in cancer research.
- Identify commonly studied cancer types, NLP tasks, data sources, and methodological trends.
- Examine an EHR-based example of NLP screening for chronic mobility disability.
- Understand the limitations of keyword-based clinical text screening.
- Compare the broad development from rule-based methods to machine learning, deep learning, and transformer approaches.
- Identify challenges involving data quality, privacy, explainability, generalisability, and clinical integration.
- Propose a practical framework for future AI-enhanced, context-aware EHR analytics.

3.2 Research Questions

RQ1: How is NLP being applied to EHRs and clinical notes in cancer research?

RQ2: Which cancer types, clinical tasks, and NLP approaches are commonly represented in the reviewed literature?

RQ3: How can NLP-based screening help identify possible chronic mobility disability from clinical notes?

RQ4: Why is manual contextual validation necessary after keyword-based screening?

RQ5: What technical and clinical challenges must be addressed before broader use of EHR-based NLP systems?

3.3 Scope of the Study

The systematic review covers peer-reviewed English-language journal research from 2019 to 2024 and uses Scopus as the search database. The review focuses on research that directly combines cancer-related applications, EHRs or clinical text, and NLP. The mobility example focuses on colorectal cancer patients and wheelchair-related information in clinical notes[5].

3.4 Important Methodological Position

The paper should be understood as an evidence synthesis and methodological framework. It does not claim that a new predictive NLP model was trained and experimentally evaluated. The reported mobility numbers come from the described EHR screening and manual validation evidence[6]. Therefore, new accuracy, precision, recall, or F1 values should not be reported unless a future study creates labelled data and conducts a formal model evaluation.

3.5 Overall Research Logic

The research follows a simple logic: understand what existing cancer-NLP research has achieved; examine a practical example of extracting mobility information from clinical notes; identify where simple extraction fails; and use these observations to design a stronger future framework. This keeps the paper useful while avoiding unsupported claims.

4. RESEARCH METHODOLOGY

4.1 Overall Design

The study has two connected components. Part A is a systematic literature review of NLP applications in cancer research. Part B is an EHR-based mobility screening example followed by manual contextual validation. The two components are related conceptually, but they do not represent one combined dataset.

4.2 Part A: Systematic Literature Review

Scopus was selected as the main literature database. The search used the concepts Electronic Health Records, EHRs, Clinical Notes, Cancer, Natural Language Processing, and NLP. These concepts were searched in titles, abstracts, and keywords. The purpose was to identify research directly related to NLP-based analysis of clinical records in cancer research[5-7]

4.3 Inclusion Criteria

- Peer-reviewed journal article.
- Published from 2019 onward and within the defined review period.
- Written in English.
- Focused on cancer research or cancer-related healthcare.
- Used EHRs, clinical notes, or related clinical text.
- Applied at least one NLP method to clinical information[6].
- Reported original research findings rather than only reviewing previous work.

4.4 Exclusion Criteria

- Conference papers and book chapters.
- Review articles, editorials, and opinion papers.
- Non-English publications.
- Studies published before 2019.
- Research without EHR or clinical-note data.
- Studies without an NLP application.
- Work discussing cancer and NLP without directly combining NLP with EHR or clinical-note cancer research.

4.5 Data Extraction

For each eligible paper, information was collected on publication details, cancer type, research objective, NLP task, NLP method or model, dataset, patient or record count, clinical information analysed, reported results, challenges, limitations, and future directions. When available, performance measures such as accuracy, precision, recall, and F1-score were recorded.

4.6 Study Taxonomy

The selected studies were organised by cancer type, data source, NLP task, technical approach, and research purpose. Main task groups included information extraction, text classification, named entity recognition, clinical concept identification, prediction, recurrence detection, prognosis, and outcome analysis. Technical groups included rule-based methods, traditional machine learning, deep learning, and transformer-based models.

5. EHR-BASED CHRONIC MOBILITY DISABILITY ANALYSIS

5.1 Data Source

The mobility example used the Research Patient Data Registry (RPDR), described as a large healthcare repository containing information from Partners Healthcare providers in eastern Massachusetts[6]. The source contained more than 2 billion clinical facts related to approximately 6.5 million individuals. Available information included demographics, diagnoses, procedures, laboratory results, medications, EHR information, and clinical notes.

5.2 Study Population

The cohort consisted of colorectal cancer patients aged 21–75 years with a new colorectal cancer diagnosis between 2005 and 2017 and sufficient EHR information. ICD-9-CM and ICD-10-CM codes were used to identify the cancer cohort. A total of 14,877 patients were included, with approximately 303,182 clinical notes available. Information was collected for up to five years before the initial cancer diagnosis.

5.3 Why Mobility Information Matters

Mobility and functional limitations are not always stored in dedicated structured fields. Important information may instead appear in narrative descriptions such as difficulty walking, limited mobility, weakness, falls, inability to walk independently, wheelchair use, or need for assistance. This makes clinical text an important source for identifying possible chronic mobility disability.

5.4 Initial NLP Screening

The initial screening searched clinical notes for wheelchair-related words and phrases. The aim was not to diagnose disability automatically. Instead, the keywords acted as a first filter that reduced the very large number of notes to a smaller group that could be reviewed more closely.

5.5 Manual Contextual Validation

The screening identified 1,482 notes associated with 420 patients. Manual review then examined whether the note described the patient's own wheelchair use, why the wheelchair was needed, and how long it was used. The review also considered whether the statement was temporary, historical, planned, related to another person, or unclear.

After contextual review, 286 notes involving 105 patients contained information about both the reason and duration of wheelchair use. These 105 patients represented approximately 0.7% of the 14,877-patient cohort. The result illustrates the difference between detecting a word and identifying a clinically meaningful condition.

6. STUDY SELECTION AND PRISMA-STYLE RESULTS

6.1 Study Selection

Screening stage	Records
Scopus records identified	272
After duplicate removal	255
After title/abstract screening	130
Full-text articles retrieved	117
Final included studies	94

The initial Scopus search identified 272 records. Seventeen duplicate records were removed, leaving 255 records for title and abstract screening. After this stage, 130 papers were considered potentially relevant. Full texts were retrieved for 117 papers. The remaining 13 could not be accessed in full text. After detailed eligibility assessment, 23 articles were excluded and 94 studies were retained for qualitative synthesis.

6.2 PRISMA-Style Flow

Identification → 272 Scopus records → 17 duplicates removed → 255 records screened → 125 records excluded at screening → 130 full-text candidates → 13 full texts unavailable → 117 full texts assessed → 23 excluded after eligibility review → 94 studies included.

6.3 Interpretation of the Review Set

The final 94 studies provide the evidence base for understanding how NLP is being used in cancer-related EHR research. The studies cover different cancer types, datasets, clinical documents, NLP tasks, and evaluation methods. Because of this heterogeneity, the synthesis focuses on broad patterns and practical lessons rather than ranking individual models.

6.4 Mobility Screening Summary

Measure	Result	Approximate proportion
Colorectal cancer patients	14,877	100%
Clinical notes	303,182	100%
Notes with wheelchair-related terms	1,482	≈0.5% of notes
Patients represented by screened notes	420	≈2.8% of patients
Notes with reason + duration	286	≈19.3% of screened notes
Patients after contextual validation	105	≈0.7% of patients

The mobility values are descriptive findings from the studied EHR dataset. They should not be interpreted as population prevalence estimates because the cohort and data source were specific to the analysed healthcare records.

7. RESULTS: NLP APPLICATIONS IN CANCER RESEARCH

7.1 Commonly Studied Cancer Types

Breast, lung, and colorectal cancers were among the commonly studied cancer areas in the reviewed literature. NLP was applied to clinical notes, pathology reports, radiology reports, treatment records, and other EHR information. The availability of large clinical datasets and the need to extract information from complex documentation make these areas suitable for NLP research.

7.2 Common NLP Tasks

- Information extraction from clinical notes.
- Text classification of documents or patient records.
- Named entity recognition for diseases, symptoms, medications, and procedures.
- Clinical concept identification and normalization.
- Patient cohort identification.
- Cancer staging and diagnosis extraction.
- Treatment and medication extraction.
- Recurrence and prognosis analysis.
- Patient outcome analysis.

7.3 Cancer Diagnosis and Patient Identification

NLP can help researchers identify patients who meet particular cancer-related criteria. This is useful when creating research cohorts or finding records for clinical studies. However, a cancer term can describe a current diagnosis, a previous diagnosis, a suspected condition, or family history. Context is therefore necessary even when the target concept appears clearly in the text.

7.4 Symptoms and Clinical Findings

Clinical notes often contain symptoms and observations that are not represented as structured fields. NLP can identify variations in wording and group related expressions. This can support research into symptoms, functional status, disease complications, and patient outcomes.

7.5 Treatment, Staging, and Prognosis

Cancer treatment information may include surgery, chemotherapy, radiotherapy, medication changes, treatment response, and treatment discontinuation. Disease progression may be described through stage, recurrence, metastasis, or follow-up observations. NLP can help extract these details from narrative records and support large-scale outcome research.

7.6 Main Finding

The central result from the literature review is that NLP is useful because important cancer information is often distributed across free-text documents. The value of NLP is not simply that it reads text faster; it helps turn narrative information into a form that can be searched, compared, and analysed at research scale.

8. RESULTS: DEVELOPMENT OF NLP APPROACHES

8.1 Rule-Based and Keyword Approaches

Earlier and simpler systems often used predefined keywords, dictionaries, patterns, and rules. These approaches have an important advantage: they are relatively easy to understand and can be implemented quickly. They are also useful for initial screening when the target terms are known.

Their main weakness is context. A keyword can be present without representing the target condition. A system searching for 'wheelchair' cannot automatically distinguish current chronic use from temporary use, planned use, or mention of another person.

8.2 Machine Learning

Traditional machine learning approaches learn patterns from labelled examples. They can use text features to classify records or identify clinical concepts. Their performance depends heavily on the quality and representativeness of the labelled data. They can also be affected by differences in documentation style between institutions.

8.3 Deep Learning

Deep learning methods can represent more complex patterns in clinical language. Models such as neural sequence models can learn relationships that are difficult to encode through manual rules. However, they generally require sufficient data, computational resources, and careful evaluation.

8.4 Transformer-Based Models

Transformer models have become important because they can capture relationships across words and sentences and can be adapted to clinical tasks. The reviewed cancer examples include transformer-based approaches such as RoBERTa and GatorTron. Their reported performance in individual studies can be strong, but performance values should always be interpreted within the specific dataset and task.

8.5 Practical Interpretation

The progression from rules to transformers should not be viewed as a simple replacement of one method by another. A practical EHR analytics system may use several levels: rules for initial filtering, machine learning or transformers for context-aware classification, and human review for difficult cases. The mobility example strongly supports this layered approach.

8.6 Need for Fair Comparison

The included studies used different populations, datasets, note types, annotation methods, cancer types, and evaluation metrics.

Therefore, the reported accuracy or F1 values cannot be treated as a common benchmark. A fair future comparison should use the same labelled dataset, the same data split, and consistent evaluation procedures.

9. RESULTS: MOBILITY SCREENING AND VALIDATION

9.1 Initial Screening

The mobility analysis began with approximately 303,182 clinical notes belonging to 14,877 colorectal cancer patients. Searching for wheelchair-related terms reduced this large collection to 1,482 potentially relevant notes linked to 420 patients. This represents approximately 0.5% of the notes and 2.8% of the patients.

9.2 What the Screening Did Well

The main benefit was reduction of manual workload. Instead of reviewing every clinical note, researchers could focus attention on a much smaller set of records. This is a realistic and useful role for NLP in research: it can act as a screening assistant that finds potentially relevant information at scale.

9.3 Why Screening Was Not Enough

The presence of a wheelchair-related term did not prove chronic mobility disability. The term could refer to temporary use after surgery or injury, another person, a future plan, equipment arrangements, or a statement without enough information to determine duration. Therefore, the 1,482 notes should be understood as candidate records rather than confirmed cases.

9.4 Contextual Review

Manual review considered three key questions: Did the statement describe the patient? Why was the wheelchair needed? How long was it used? The reviewers also considered time, history, future plans, and whether the information was clinically sufficient.

9.5 Validated Findings

After review, 286 notes involving 105 patients contained information about both reason and duration. The reduction from 1,482 keyword hits to 286 contextually relevant notes is a practical demonstration of the limits of simple keyword matching.

9.6 Key Lesson

The most important lesson is that extraction and interpretation are different. NLP can identify where relevant language appears, but the meaning of that language depends on context. Future systems should therefore use context-aware classification and temporal reasoning instead of relying only on isolated words.

9.7 Evaluation Caution

Because the current study does not report a newly trained model with a formal labelled test set, no new accuracy, precision, recall, or F1 score is claimed. The observed counts are screening and validation findings, not model-performance metrics.

10. PROPOSED AI-ENHANCED EHR ANALYTICS FRAMEWORK

The findings support a practical end-to-end framework for future EHR analytics. The framework connects structured EHR information with unstructured clinical text and keeps human validation as an important safety and quality step.

Stage	Main activity	Purpose
1. Problem definition	Define research question and target clinical concepts	Keep the task specific and measurable
2. EHR data collection	Collect eligible patient records and notes	Create the research dataset
3. Cleaning and de-identification	Remove duplicates and protect identity	Improve data quality and privacy
4. Data separation	Separate structured fields from free text	Apply suitable processing to each source
5. Text preprocessing	Normalize notes, abbreviations, and formatting where appropriate	Prepare clinical text

6. Entity/information extraction	Identify cancer, symptoms, treatments, mobility concepts	Convert text into usable information
7. Context and temporal analysis	Identify negation, patient reference, time, history, and planned events	Reduce false positives
8. Classification/cohort identification	Assign clinically meaningful categories	Build research cohorts
9. Human validation	Review uncertain or high-value cases	Confirm clinical meaning
10. Evaluation	Compare predictions with labelled reference data	Measure reliability
11. Interpretation	Analyse results, errors, and clinical relevance	Support research conclusions
12. Research output	Report findings, limitations, and future work	Produce transparent evidence

10.1 Mobility-Specific Flow

Eligible colorectal cancer cohort → clinical notes → wheelchair/mobility screening → context review → reason and duration extraction → manual validation → relevant cases → error analysis → future model development.

10.2 Cancer Research Flow

Research question → Scopus search → duplicate removal → title/abstract screening → full-text assessment → 94 included studies → structured extraction → cancer/NLP taxonomy → trend analysis → challenges and future directions.

10.3 Important Separation

These workflows should not be presented as if they were one combined dataset. The systematic review uses published research papers. The mobility example uses the described EHR cohort and clinical notes. The overall framework is a proposed methodological connection between evidence synthesis, EHR extraction, validation, and future model development.

11. FUTURE MODEL DEVELOPMENT AND EVALUATION

11.1 Reference Dataset

A future predictive model should begin with a high-quality labelled dataset. Clinical notes should be reviewed by trained annotators using clear definitions. Suggested labels for wheelchair-related notes include chronic wheelchair use, temporary wheelchair use, wheelchair mentioned for another person, planned or future wheelchair use, and no relevant wheelchair information.

11.2 Training, Validation, and Test Sets

The labelled data should be divided into separate training, validation, and test sets. The training set is used to learn model parameters. The validation set is used for model selection and tuning. The test set should remain separate until final evaluation so that performance is measured on unseen records.

11.3 Evaluation Measures

Measure	Meaning
Precision	Among records predicted as positive, how many are actually positive.
Recall	Among actual positive records, how many are identified.
F1 score	Balances precision and recall.
Sensitivity	Measures the ability to identify positive cases.

Specificity	Measures the ability to identify negative cases.
AUROC	Summarises discrimination across classification thresholds.
AUPRC	Useful when positive cases are relatively uncommon.

Precision = TP/(TP+FP). Recall = TP/(TP+FN). F1 = 2 × (Precision × Recall)/(Precision + Recall). Multiple measures are preferable to accuracy alone, especially when positive cases are uncommon.

11.4 Future Model Comparison

Approach	Role	Suggested evaluation
Keyword/rule baseline	Simple initial screening	Precision, Recall, F1
Traditional machine learning	Text classification using engineered features	Precision, Recall, F1
CNN/BiLSTM	Deep learning text classification	Precision, Recall, F1
Clinical transformer	Context-aware clinical language understanding	Precision, Recall, F1
Proposed context-aware transformer	Mobility classification with contextual and temporal cues	Precision, Recall, F1, Sensitivity, Specificity

The table describes a future experimental plan, not completed experiments. Numerical model-performance results should be added only after actual training and testing.

12. DISCUSSION

The combined findings show that NLP is useful because important clinical information is often stored in free text. Structured EHR fields are easier to process, but narrative notes can contain information about symptoms, functional limitations, treatment response, disease progression, and patient circumstances that may not appear in coded fields.

In cancer research, NLP can reduce the effort needed to locate relevant patients and extract cancer-related information. The 94-study review shows applications ranging from diagnosis and patient identification to treatment, staging, recurrence, prognosis, and outcomes. The literature also shows a gradual move from keyword and rule-based methods toward machine learning, deep learning, and transformer-based approaches.

12.1 Information Extraction Versus Interpretation

The mobility example makes an important distinction. A system can extract the word 'wheelchair' very efficiently, but that does not mean it has understood the patient's condition. The same word may describe chronic use, temporary use, another person, a future plan, or a general discussion. Clinical NLP therefore needs to understand who, what, when, and why.

12.2 Human-in-the-Loop NLP

A human-in-the-loop approach is practical for healthcare research. Automated NLP can screen thousands of records and identify likely cases. Trained reviewers can then examine ambiguous or high-value cases. This approach can reduce workload while preserving human judgement where the clinical meaning is uncertain.

12.3 Validated Data as Future Training Material

The manually reviewed records can become valuable training material for future models. Examples of true chronic disability, temporary wheelchair use, unrelated mentions, and planned use can help a model learn distinctions that keyword rules cannot capture. The quality of this future model will depend on consistent annotation and representative data.

12.4 Combining Structured and Unstructured Data

A stronger system should combine diagnosis codes, laboratory results, medications, procedures, and demographic information with clinical notes, pathology reports, radiology reports, discharge summaries, and progress notes. This multimodal EHR view can provide a more complete picture of the patient's condition.

12.5 External Validation

A model that works well in one institution may perform differently elsewhere. Hospitals use different EHR systems, abbreviations, documentation practices, and patient populations. External validation across multiple healthcare settings is therefore essential before a model is considered broadly reliable.13. Limitations

13.1 Database Coverage

The systematic review used Scopus as the main database. Scopus provides broad academic coverage, but it does not contain every relevant publication. Future reviews could combine Scopus with databases such as PubMed/MEDLINE, Web of Science, and IEEE Xplore to broaden coverage.

13.2 Review Period

The review period was limited to 2019–2024. This keeps the synthesis focused on recent developments but excludes earlier foundational work and research published after the defined period.

13.3 Heterogeneity of Included Studies

The 94 studies differed in cancer type, patient population, dataset size, clinical setting, NLP task, annotation process, model, and evaluation method. Because of these differences, direct numerical comparison between studies is difficult. The findings should therefore be understood as a synthesis of trends rather than a ranking of models.

13.4 No Newly Trained Predictive Model

The present work does not report a newly trained predictive model with a formal experimental test set. The mobility analysis is based on keyword screening and manual contextual validation. Therefore, the paper does not claim new accuracy, precision, recall, or F1 values for a proposed model.

13.5 Keyword Screening Limitations

Keyword screening can create false positives when a term is present but does not describe chronic disability. It can also create false negatives when the clinical note uses another expression, such as difficulty walking, impaired ambulation, or limited functional mobility, without mentioning a wheelchair.

13.6 Incomplete Documentation

Clinical notes may not specify the start date, severity, cause, or duration of a mobility problem. Missing information affects both automated classification and human review.

13.7 Privacy and Access

EHR data contains sensitive information and is subject to ethical, security, privacy, and institutional requirements. These restrictions can make it difficult to build large multi-institutional datasets and perform external validation.

13.8 Generalisability

Differences in documentation style, terminology, EHR systems, and patient populations can reduce performance when an NLP system is moved from one institution to another. External validation is therefore a major requirement for future work.

14. FUTURE RESEARCH DIRECTIONS

14.1 Context-Aware Transformer Models

Future work should investigate clinical transformer models that can consider the surrounding text. The model should distinguish chronic from temporary conditions, patient from non-patient references, current from historical conditions, confirmed from suspected conditions, and actual from planned wheelchair use.

14.2 Expert-Labeled Datasets

High-quality labels are essential. Clinical experts or trained annotators should use clear guidelines, and agreement between reviewers should be assessed. Multi-institutional datasets would be particularly useful because they expose the model to different writing styles and patient populations.

14.3 Structured + Unstructured EHR Integration

Future systems should combine diagnosis and procedure codes, medications, laboratory results, and demographic information with clinical notes, pathology reports, radiology reports, discharge summaries, and patient histories. Combining these sources can reduce dependence on any single documentation type.

14.4 Temporal NLP

Temporal analysis is especially important for mobility disability. The system should identify when a condition started, whether it is ongoing, how long it lasted, and whether it improved or worsened. This can help distinguish chronic disability from short-term events.

14.5 Explainable NLP

An EHR NLP system should ideally show the evidence supporting an extracted result. For example, if the system identifies chronic mobility limitation, it should highlight the relevant sentence or phrase. This makes results easier to review and helps identify errors.

14.6 Privacy-Preserving NLP

Future research can examine privacy-preserving approaches such as federated learning and other secure learning methods. These approaches may support collaboration across institutions without requiring all raw patient data to be transferred to one central location. Privacy and governance requirements must still be carefully assessed.

14.7 Standardised Evaluation

Future studies should report accuracy, precision, recall, F1 score, sensitivity, specificity, and suitable AUROC or AUPRC measures. The test set should remain separate from training and tuning. External validation should be included whenever possible.

14.8 Understudied Conditions and Clinical Integration

Future research should extend beyond commonly studied cancers and investigate less-studied conditions and diverse populations. Systems should also be designed for real research and clinical workflows. The goal should be to assist professionals rather than replace human judgement.

15. INTEGRATED RESEARCH WORKFLOW

15.1 Overall Workflow

Problem definition → EHR data collection → cleaning and de-identification → structured/unstructured separation → clinical note preprocessing → NLP entity and information extraction → cancer information extraction + mobility information extraction → context and temporal analysis → classification/cohort identification → manual validation → evaluation → clinical interpretation → research output.

15.2 Systematic Review Workflow

Scopus search → duplicate removal → title and abstract screening → full-text retrieval → full-text assessment → inclusion/exclusion decisions → data extraction → cancer-type and NLP-task categorisation → trend analysis → challenges and future directions.

15.3 Mobility Screening Workflow

Eligible colorectal cancer cohort → clinical-note extraction → wheelchair keyword screening → context review → reason and duration extraction → manual validation → relevant-case identification → error analysis.

15.4 Why the Three Workflows Matter

The literature-review workflow explains what has already been done. The mobility workflow demonstrates how unstructured clinical information can be screened in a real EHR setting. The overall framework connects these lessons to future AI-enhanced EHR analytics. Keeping the three workflows conceptually separate prevents overclaiming and makes the methodology easier to reproduce.

15.5 Proposed Error Analysis

Error type	Example	Future improvement
False positive	Wheelchair mentioned for another person	Entity and patient-reference detection
Temporary use	Wheelchair after surgery	Temporal and duration analysis
Future plan	Wheelchair recommended for later use	Future-event detection
Insufficient context	Wheelchair mentioned without reason or duration	Context-aware classification
False negative	Mobility limitation described without wheelchair term	Broader concept and synonym extraction
Documentation variation	Different abbreviations or local terminology	Institution-specific adaptation and external validation

15.6 Practical Research Principle

The proposed framework treats NLP as an assistant for clinical information discovery. Automation should reduce repetitive searching, while validation and transparent evaluation should protect against incorrect interpretation. This balance is particularly important when extracted information may later be used for research cohort construction or clinical decision support.

16. LATEST RESULT ANALYSIS AND UPDATED EVIDENCE

Scope and evidence note. This section adds an analysis of the latest results reported in the supplied manuscript. It does not introduce new model-training results. The source manuscript explicitly states that the work is an evidence synthesis and a methodological framework, and that no newly trained predictive model or new model-performance scores are claimed (Source manuscript, Abstract and Sections 3.4, 9.7, and 13.4).

16.1 Systematic Literature Review Results

The Scopus search identified 272 records. After duplicate removal, title/abstract screening, full-text retrieval, and eligibility assessment, 94 studies were included in the qualitative synthesis. The included literature commonly represented breast, lung, and colorectal cancer. Information extraction and text classification were frequent tasks, while methods ranged from keyword/rule-based systems to traditional machine learning, deep learning, and transformer-based approaches. These findings describe the review set and should not be interpreted as a pooled model-performance comparison.

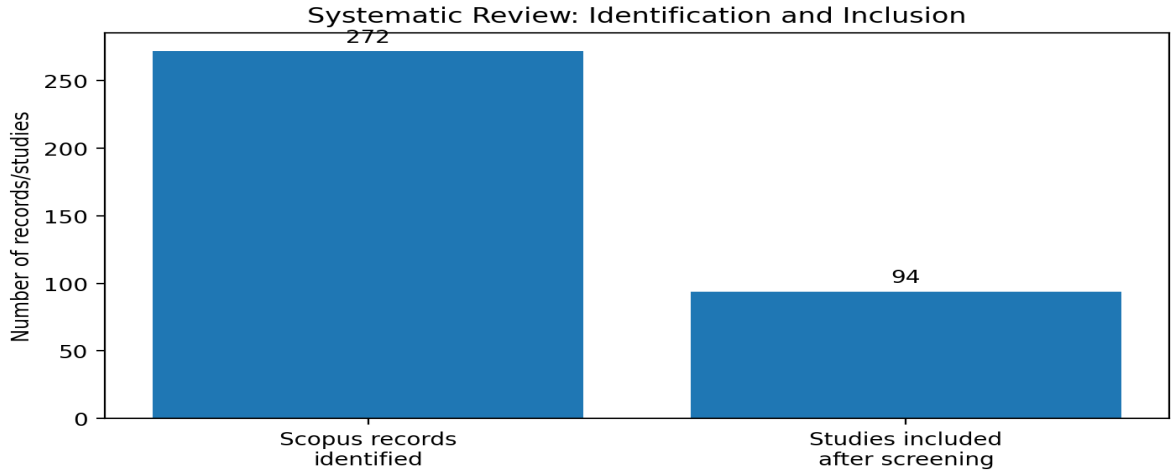


Figure 1. Initial Scopus records compared with studies included in the review.

16.2 Mobility Screening Results

The EHR example included 14,877 colorectal cancer patients and approximately 303,182 clinical notes. Keyword screening identified 1,482 wheelchair-related notes associated with 420 patients. Manual contextual review found that 286 notes, representing 105 patients, documented both the reason for wheelchair use and its duration. This is a descriptive screening and validation result, not a measure of a newly trained model's accuracy, precision, recall, or F1 score. The pattern is consistent with the published study by Agaronnik and colleagues [1].

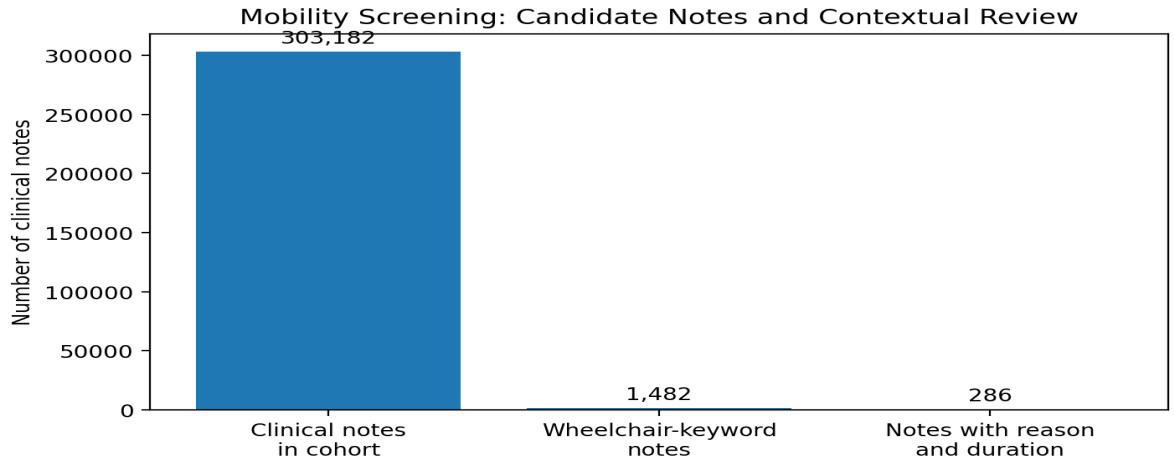


Figure 2. Clinical-note counts across the mobility-screening workflow.

16.3 Interpretation of Results

NLP can reduce the volume of clinical text requiring manual inspection and can support research-cohort construction [2, 3].Keyword hits should be treated as candidate records rather than confirmed chronic mobility-disability cases because context, patient reference, negation, temporality, and duration affect interpretation [1].

Cancer NLP studies use heterogeneous datasets, tasks, annotation procedures, and evaluation metrics; reported scores from separate studies should not be directly ranked as if they were obtained under one experimental protocol [2, 3].

The current paper is an evidence synthesis and methodological framework. It does not report a newly trained predictive model or new model-performance scores. A future model should be evaluated on expert-labelled data with a held-out test set and measures such as precision, recall, F1, sensitivity, specificity, AUROC, and AUPRC.

16.4 Results Summary

Measure	Result	Interpretation
Scopus records identified	272	Initial literature search
Studies included	94	Qualitative synthesis, 2019–2024
Colorectal cancer patients	14,877	EHR mobility cohort
Clinical notes	303,182	Notes associated with cohort
Wheelchair-keyword notes	1,482; 420 patients	Candidate notes after keyword screening
Notes with reason and duration	286; 105 patients	Contextually relevant documentation

Natural Resources for Human Health
eISSN: 2583-1194

16.5 Latest reported quantitative results

Analysis stage	Reported result	Derived proportion	Interpretation
Systematic review	272 records initially identified	94/272 = 34.6% included	94 studies formed the final qualitative synthesis.
After duplicate removal	255 records screened	17/272 = 6.25% duplicates	Seventeen duplicate records were removed.
Full-text assessment	117 full texts assessed	94/117 = 80.3% included	23 articles were excluded after eligibility review.
EHR cohort	14,877 colorectal cancer patients	100% of cohort	Cohort-specific results; not a population prevalence estimate.
Clinical-note collection	≈303,182 notes	100% of notes	Starting volume for wheelchair-related screening.
Keyword screening	1,482 notes; 420 patients	≈0.49% of notes; ≈2.82% of patients	Candidate records requiring contextual review.
Reason + duration documented	286 notes	≈19.3% of keyword-screened notes	Notes containing both reason and duration information.
Contextually validated patients	105 patients	≈0.71% of cohort; 25.0% of 420 screened patients	Patients represented by notes with both reason and duration information.

16.6 Visual analysis of mobility-screening results

Figure 1 compares the reported clinical note counts across the screening funnel. A logarithmic vertical axis is used because the starting collection (≈303,182 notes) is much larger than the keyword-screened set (1,482) and the contextually relevant notes (286). The counts and percentages are derived from the supplied manuscript.

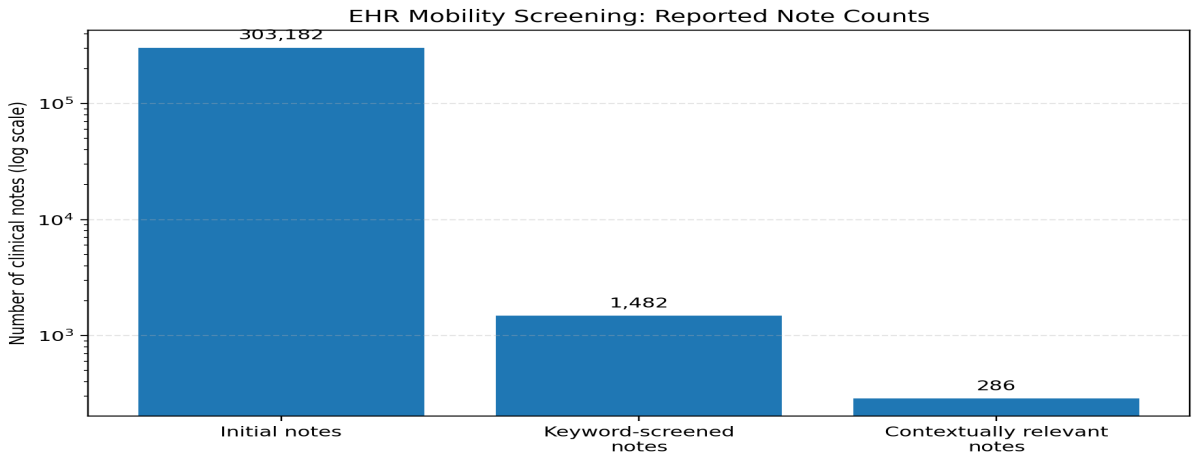


Figure 1. Reported EHR mobility-screening note counts (logarithmic scale).

16.6 Result interpretation

The literature-review funnel retained 94 of 272 initially identified records (34.6%). This is a study-selection proportion, not an estimate of the quality or effectiveness of NLP models.

Keyword screening reduced approximately 303,182 notes to 1,482 candidate notes—about 0.49% of the original note

collection. This demonstrates a substantial reduction in the number of records needing manual review.

After contextual manual review, 286 notes involving 105 patients contained information about both the reason and duration of wheelchair use. The 105 patients correspond to about 0.71% of the 14,877-patient cohort. This should not be presented as general-population prevalence because the cohort and source institution are specific.

The reduction from 1,482 keyword hits to 286 notes with reason-and-duration information shows why keyword detection is a screening step rather than a definitive classifier for chronic mobility disability. The manuscript does not provide the labelled test set needed to calculate sensitivity, specificity, precision, recall, F1, or predictive accuracy for a newly trained model.

The cancer-study performance examples already present in the manuscript (including approximately 81% accuracy for an ovarian-cancer example, approximately 94% for a cervical-cancer example, GatorTron F1 values of 0.885–0.950, a 90.4% skin-cancer matching rate, and an approximately 83% CNS result with AUROC 0.92) come from different studies and tasks. They are not directly comparable and should be retained as literature examples, not combined into a single leaderboard.

17. OVERALL FINDINGS AND IMPLICATIONS

17.1 Main Findings

- The systematic review identified 94 eligible studies from an initial 272 Scopus records.
- Breast, lung, and colorectal cancer were among the commonly represented cancer areas.
- Information extraction and text classification were frequently used NLP tasks.
- The literature showed movement from keyword/rule-based approaches toward machine learning, deep learning, and transformer methods.
- The mobility example included 14,877 colorectal cancer patients and about 303,182 clinical notes.
- Keyword screening identified 1,482 notes from 420 patients.
- Manual contextual review identified 286 notes involving 105 patients with information about both wheelchair-use reason and duration.
- Keyword presence alone was not sufficient to confirm chronic mobility disability.
- Future systems require labelled data, context awareness, temporal reasoning, explainability, privacy protection, and external validation.

17.2 Research Implications

For cancer research, NLP can make large clinical datasets more usable by extracting information that is otherwise difficult to locate in free text. It can support cohort construction, symptom analysis, treatment research, staging, recurrence studies, prognosis, and outcome analysis.

For mobility research, the main implication is that functional disability information can be present in narrative clinical notes even when it is not obvious in structured data. NLP can help locate these records, but meaningful identification requires context and validation.

17.3 Methodological Implications

The findings support a layered approach rather than a single automated method. A practical pipeline may begin with efficient rules or keyword screening, move to context-aware machine learning or transformer classification, and use human review for uncertain cases. This approach is realistic because it combines computational scale with clinical interpretation.

17.4 Clinical and Research Relevance

The proposed framework can help researchers organize information from large EHR collections without claiming that automation can replace clinicians. Its value is strongest when the system makes large-scale review faster, identifies candidate records, provides evidence for its decisions, and allows human reviewers to correct or confirm difficult cases.

17.5 Responsible Use

Because EHR information is sensitive, future development must consider privacy, governance, fairness, explainability, and secure data handling. High performance on one dataset is not enough. A reliable system should be tested on unseen data and, where possible, on records from other healthcare institutions.

18. CONCLUSION

This study presents a concise framework for AI-enhanced NLP analysis of Electronic Health Records, with a focus on cancer research and chronic mobility disability identification. The systematic literature review included 94 studies published during 2019–2024 and showed that NLP is increasingly used to extract and analyse clinical information from EHRs and clinical notes.

The reviewed research covered applications such as cancer diagnosis, patient identification, symptom extraction, treatment analysis, staging, prognosis, recurrence detection, and patient outcome analysis. The literature also showed a gradual shift from keyword and rule-based methods toward machine learning, deep learning, and transformer-based approaches. This shift reflects the need for systems that can understand clinical context rather than only detect individual words.

The mobility screening example provided an important practical lesson. From approximately 303,182 clinical notes belonging to 14,877 colorectal cancer patients, keyword screening identified 1,482 notes associated with 420 patients. After manual contextual review, 286 notes involving 105 patients contained information about both the reason and duration of wheelchair use. The difference between these numbers demonstrates why keyword detection should be treated as screening rather than final classification.

The study therefore supports a human-in-the-loop and context-aware approach. Future work should create expert-labelled clinical datasets and train models that can recognise negation, time, patient reference, temporary conditions, historical information, and planned future events. Structured EHR information should be combined with unstructured clinical text whenever possible.

Future model development should also use separate training, validation, and test sets and report appropriate measures such as precision, recall, F1 score, sensitivity, specificity, AUROC, and AUPRC. External validation across healthcare institutions will be important because documentation styles, EHR systems, terminology, and patient populations can differ.

Overall, the findings show that AI-enhanced NLP can make large-scale EHR analysis more efficient and can help researchers find clinically important information that may be hidden in free text. At the same time, reliable clinical NLP requires careful context interpretation, human validation, privacy protection, explainability, and rigorous evaluation. The proposed framework provides a practical foundation for future research into intelligent EHR analytics for cancer research and chronic mobility disability identification.

Selected Study Tables

Item	Key evidence
Review evidence	94 included studies from 272 initial Scopus records
Cancer focus	Breast, lung, colorectal among commonly studied cancers
Frequent tasks	Information extraction and text classification
Mobility cohort	14,877 colorectal cancer patients
Clinical notes	≈303,182
Keyword screening	1,482 notes; 420 patients
Contextually relevant	286 notes; 105 patients
Main limitation	Keyword presence does not equal chronic disability
Future priority	Context-aware, explainable, privacy-preserving and externally validated NLP

Final Statement

The central message of this research is simple: NLP can help find important information in large clinical records, but finding a word is not the same as understanding a clinical condition. The most useful future systems will combine AI-based extraction, contextual reasoning, structured and unstructured EHR data, and appropriate human validation.

Research Contribution

- Connects cancer-focused EHR NLP research with a practical mobility-disability screening example.
- Summarises 94 studies and identifies common cancer areas, NLP tasks, methods, and challenges.
- Shows why automated screening and human validation should work together.

Provides a clear framework for future context-aware NLP model development without overstating current evidence.

REFERENCES

[1] N. D. Agaronnik, C. Lindvall, A. El-Jawahri, W. He, and L. I. Iezzoni, "Challenges of Developing a Natural Language Processing Method With Electronic Health Records to Identify Persons With Chronic Mobility Disability," Archives of Physical Medicine and Rehabilitation, vol. 101, no. 10, pp. 1739–1746, 2020. doi: 10.1016/j.apmr.2020.04.009. Available: https://pubmed.ncbi.nlm.nih.gov/32446905/

[2] L. Wang et al., "Assessment of Electronic Health Record for Cancer Research and Patient Care Through a Scoping Review of Cancer Natural Language Processing," JCO Clinical Cancer Informatics, vol. 6, 2022. doi: 10.1200/CCI.22.00006. Available: https://pubmed.ncbi.nlm.nih.gov/35917480/

[3] J.-A. Sim, X. Huang, M. R. Horan, J. N. Baker, and I.-C. Huang, "Using natural language processing to analyze unstructured patient-reported outcomes data derived from electronic health records for cancer populations: a systematic review," Expert Review of Pharmacoeconomics & Outcomes Research, vol. 24, no. 4, pp. 467–475, 2024. doi: 10.1080/14737167.2024.2322664. Available: https://pubmed.ncbi.nlm.nih.gov/38383308/

[4] B. Kayira, H. R. A. Elyazori, K. Lybarger, F. M. Walter, C. Chelala, and G. Funston, "Natural Language Processing of Clinical Notes for Cancer Research and Patient Care Prior to Widespread Adoption of Generative AI: Scoping Review," JMIR AI, vol. 5, e73481, 2026. doi: 10.2196/73481. Available: https://pubmed.ncbi.nlm.nih.gov/42133609/

[5] "Performance of Natural Language Processing for Information Extraction From Electronic Health Records Within Cancer: Systematic Review," JMIR Medical Informatics, 2025. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12431712/

[6] M. Bilal, A. Hamza, and N. Malik, "NLP for analyzing electronic health records and clinical notes in cancer research: A review," Journal of Pain and Symptom Management, vol. 69, no. 5, pp. e374–e394, 2025, doi: 10.1016/j.jpainsymman.2025.01.019.

[7] S. Dahl, M. Bøgsted, T. Sagi, and C. Vesteghem, "Performance of natural language processing for information extraction from electronic health records within cancer: Systematic review," JMIR Medical Informatics, vol. 13, Art. no. e68707, 2025, doi: 10.2196/68707.

[8] J. A. Sim et al., "Using natural language processing to analyze unstructured patient-reported outcomes data derived from electronic health records for cancer populations: A systematic review," Expert Review of Pharmacoeconomics & Outcomes Research, vol. 24, no. 4, pp. 467–475, 2024, doi: 10.1080/14737167.2024.2322664.

[9] M. Gholipour, R. Khajouei, P. Amiri, S. Hajesmaeel Gohari, and L. Ahmadian, "Extracting cancer concepts from clinical notes using natural language processing: A systematic review," BMC Bioinformatics, vol. 24, Art. no. 405, 2023, doi: 10.1186/s12859-023-05480-0.

[10] Y. Wieland-Jorna et al., "Natural language processing systems for extracting information from electronic health records about activities of daily living: A systematic review," JAMIA Open, vol. 7, no. 2, Art. no. ooae044, 2024, doi: 10.1093/jamiaopen/ooae044.

[11] K. Kreimeyer et al., "Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review," Journal of Biomedical Informatics, vol. 73, pp. 14–29, 2017, doi: 10.1016/j.jbi.2017.07.012.

[12] J. Lee et al., "BioBERT: A pre-trained biomedical language representation model for biomedical text mining," Bioinformatics, vol. 36, no. 4, pp. 1234–1240, 2020, doi: 10.1093/bioinformatics/btz682.

[13] Jadeja, Vijaysinh, et al. "A Robust and Secure High-Capacity Coverless Information Hiding Scheme Using Deep Learning." International Journal of Computer Information Systems & Industrial Management Applications 18 (2026): 629.

[14] Jadeja, VK. "Advancements In High-Capacity Coverless Information Hiding: A Comprehensive Review Of Robust Hashing Techniques." African Journal Of Biomedical Research (2024): 706-714.
[15] Jadeja, Vijaysinh, et al. "High-Capacity Coverless Information Hiding Based On Robust and Deep Learning." International Journal of Artificial Intelligence and Machine Learning 6.11s (2026): 1647-1658.
[16] M. J. Page et al., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," BMJ, vol. 372, Art. no. n71, 2021, doi: 10.1136/bmj.n71.