

Intelligent EHR Analytics Using AI-Enhanced NLP for Cancer Research and Chronic Mobility Disability Identification

Jagrati Shekhawat¹ and Viral Parekh²

¹Research Scholar, Monark University, Computer Engineering Gujrat, India

²Associate Professor, Department of Computer Engineering, Monark University.

Received: 14th June, 2026; Revised: 20th July 2026; Accepted: 20th August, 2026; Available Online: 10th September 2026

ABSTRACT

Electronic Health Records (EHRs) contain important information about patients, including diagnoses, symptoms, laboratory results, treatments, clinical observations, and health outcomes. Although some EHR information is stored in structured formats, a large amount is contained in unstructured clinical notes. These notes are difficult to analyze manually because they may include abbreviations, incomplete sentences, spelling variations, negation, temporal expressions, and ambiguous clinical context.

Natural Language Processing (NLP), supported by Artificial Intelligence (AI), provides methods for converting unstructured clinical text into useful structured information. This paper examines the use of NLP in two major healthcare applications: cancer research and identification of chronic mobility disabilities. It also presents a machine-learning experiment for Chronic Kidney Disease (CKD) prediction. The cancer-related analysis focuses on information extraction, text classification, clinical phenotyping, treatment analysis, and outcome prediction. For mobility-disability identification, NLP was used to screen 303,182 clinical notes from 14,877 colorectal cancer patients for wheelchair-related terms. The method identified 1,482 notes associated with 420 patients. Manual validation found that only 286 notes involving 105 patients contained information about both the reason and duration of wheelchair use.

In the CKD experiment, Logistic Regression, Random Forest, and Gradient Boosting models were evaluated using 500 records. Logistic Regression achieved the best performance, with 98.0% accuracy, 96.15% precision, 100% recall, and 98.04% F1-score. These results indicate the potential of AI-based methods for healthcare-data analysis. However, keyword-based NLP requires contextual validation, and the CKD model requires external validation using larger and clinically representative datasets before clinical deployment.

Keywords: Electronic Health Records, Natural Language Processing, Artificial Intelligence, Cancer Research, Chronic Mobility Disability, Chronic Kidney Disease, Machine Learning, Clinical Text Mining.

How to cite this article: Jagrati Shekhawat¹ and Viral Parekh². Intelligent EHR Analytics Using AI-Enhanced NLP for Cancer Research and Chronic Mobility Disability Identification. *Int J Drug Deliv Technol.* 2026;16(83s):719-726. DOI: 10.25258/ijddt.16.83s.88.

Source of support: Nil.

Conflict of interest: None

I. INTRODUCTION

Electronic Health Records (EHRs) are digital systems that store patient information for healthcare delivery, clinical decision-making, administration, and research. Typical EHRs include demographic information, diagnoses, medications, laboratory results, procedures, imaging reports, pathology reports, and clinical notes [1]. These data can be divided into two broad categories: structured and unstructured data.

Structured data are stored in predefined fields. Examples include laboratory values, diagnosis codes, medication records, procedure codes, and vital signs. Because structured data follow a fixed format, they can usually be analyzed using conventional database and statistical methods. In contrast, unstructured data are usually written as free text. Examples include physician notes, nursing notes, discharge summaries, radiology reports, pathology reports, and rehabilitation documentation [1], [2]. Unstructured clinical text contains information that may not be recorded accurately in structured fields. For example, clinical notes may describe a patient's symptoms, treatment response, functional status, disease progression, recurrence, complications, and ability to walk. However, manual review of thousands or millions of clinical documents is time-consuming and expensive. It is also subject to reviewer fatigue and inconsistency.

Natural Language Processing (NLP) provides computational techniques for analyzing human language. In healthcare, clinical NLP can identify medical concepts, extract relationships, classify documents, detect symptoms, identify treatments, and support patient phenotyping [2], [3]. More recent systems use deep-learning and transformer-based models, such as BERT and RoBERTa, to capture the context in clinical language [4], [10], [11].

AI-enhanced analysis of electronic health records (EHR) has emerged as a transformative approach in modern healthcare research, enabling the extraction of clinically meaningful insights from vast repositories of structured and unstructured patient data that would otherwise remain underutilized due to their scale and complexity. This paper examines this growing field through three interconnected but distinct lines of inquiry. First, it explores natural language processing (NLP) applications in cancer research, where techniques such as named entity recognition, relation extraction, and large language model-based text mining are used to parse pathology reports, oncology notes, and radiology findings to identify tumor characteristics, staging information, treatment responses, and adverse events that are often buried in free-text narratives rather than coded fields. Second, the paper addresses the identification of chronic mobility disabilities from clinical notes, a domain where disability

status is frequently underdocumented or inconsistently coded in structured EHR fields, making NLP-driven phenotyping essential for surfacing functional impairment information critical to equitable care delivery, resource allocation, and disability-inclusive research. Third, it investigates machine-learning-based prediction of Chronic Kidney Disease (CKD), leveraging structured clinical variables such as laboratory results, comorbidities, and longitudinal trends to build predictive models capable of flagging at-risk patients earlier than traditional diagnostic pathways allow, thereby supporting timely intervention and disease management. Collectively, these three strands illustrate the versatility of AI methodologies in EHR analysis, spanning unstructured text mining, phenotype identification for underrepresented health conditions, and structured-data risk prediction, while also highlighting shared methodological challenges such as data heterogeneity, label sparsity, model interpretability, and the need for rigorous validation before clinical deployment.

II. BACKGROUND

A. Electronic Health Records

EHRs contain longitudinal information about a patient's interaction with the healthcare system. The information may be collected from hospitals, outpatient clinics, laboratories, pharmacies, imaging centers, and rehabilitation services. Electronic Health Records comprise a rich and heterogeneous set of components, each contributing a distinct layer of clinical insight essential for comprehensive patient care and downstream AI-driven analysis. Patient demographics establish the foundational context, including age, sex, ethnicity, and socioeconomic indicators, which are often critical covariates in predictive modeling and disparity research. Diagnoses and problem lists provide structured, often ICD-coded representations of a patient's medical history and active health concerns, while medication orders and prescriptions capture pharmacological interventions, dosages, and treatment timelines that inform both efficacy assessments and adverse event detection. Laboratory measurements offer quantitative, longitudinal insight into physiological function, serving as key inputs for models predicting conditions like Chronic Kidney Disease, whereas procedures and surgeries document interventional history that can signal disease severity or progression. Radiology and pathology reports, though typically unstructured, contain dense diagnostic narratives, tumor characterizations, and imaging findings that are prime targets for NLP-based extraction, particularly in cancer research contexts. Clinical progress notes and discharge summaries synthesize the ongoing narrative of a patient's care trajectory, capturing clinician reasoning, treatment responses, and care transitions that are frequently the richest yet most challenging sources for text mining. Finally, functional and social information, though often inconsistently documented, is indispensable for identifying conditions such as chronic mobility disabilities, as it reflects patient-reported outcomes, activities of daily living, and social determinants of health that structured fields alone frequently fail to capture. Together, these components form a multidimensional data ecosystem in which structured and unstructured elements must be jointly leveraged to achieve holistic, AI-enhanced

patient assessment. The structure of EHR data varies across healthcare institutions. Different hospitals may use different systems, terminology, templates, abbreviations, and documentation practices. This lack of standardization creates challenges for secondary research use.

B. Structured and Unstructured Clinical Data

Structured data are relatively easy to search and analyze. For example, a researcher can identify patients with a particular diagnosis code or laboratory value using a database query. However, structured fields may not fully represent the patient's clinical condition. While a diagnosis code can confirm that a patient has cancer, it usually cannot explain the fuller clinical picture surrounding that diagnosis. For example, the code alone does not describe what symptoms the patient is currently experiencing, why a doctor decided to change a treatment plan, or whether the disease is getting better or worse over time. It also does not show whether the patient is dealing with complications caused by treatment, such as side effects from chemotherapy. In a similar way, a code cannot tell us whether a patient needs a wheelchair, or whether that need is just temporary, like recovering from surgery, or a long-term, chronic condition. This kind of detailed, real-world information is usually written down by doctors and nurses in free-text clinical notes rather than stored in structured fields. Because of this, clinical notes are a valuable source of information that helps fill in the gaps left by diagnosis codes alone [1], [2].

C. Clinical Natural Language Processing

Clinical NLP refers to the use of computational methods to analyze medical language, and it relies on a series of connected tasks that together turn free-text notes into usable information. Tokenization and sentence segmentation break the raw text into words and sentences, providing the basic units for further analysis. Named entity recognition and medical concept extraction then identify important terms in the text, such as diseases, medications, symptoms, or procedures, and link them to standard medical vocabularies. Text classification sorts notes or sentences into meaningful categories, such as identifying the type of note or flagging a specific clinical issue. Relation extraction connects related pieces of information, for example linking a medication to the condition it treats, while negation detection determines whether a condition is actually present or explicitly ruled out, such as recognizing that "no signs of infection" means infection is absent. Temporality detection identifies when an event occurred, such as in the past, present, or future, and clinical phenotyping combines multiple pieces of extracted information to identify whether a patient fits a broader clinical pattern, such as a chronic condition. Finally, information summarization condenses long, detailed notes into shorter, more digestible summaries. For example, using these techniques, an NLP system may identify the following information from a clinical note:

Entity: "wheelchair" → recognized as a mobility-related medical concept.

Negation: "no evidence of metastasis" → correctly identified as a negated finding.

Temporality: "patient reports pain for the past three months" → identified as an ongoing, past-to-present symptom.

Relation: "prescribed metformin for diabetes" → medication linked to its associated diagnosis

Using these NLP techniques, the system is able to pull out several important pieces of information from the clinical note and organize them clearly. It identifies the assistive device mentioned, in this case a wheelchair, and connects it to the underlying functional condition causing the need for it, which here is lower-extremity weakness. It also picks up on how long the patient has been dealing with this issue, noting a duration of eight months, and determines the temporal status as ongoing, meaning the condition is still present rather than resolved. Bringing these details together, the system can then make a reasonable judgment that this may represent a possible chronic disability, since a mobility issue lasting eight months and still ongoing suggests a long-term rather than temporary condition. This kind of structured extraction from unstructured text shows how clinical NLP can surface meaningful, actionable insights that would otherwise remain hidden within free-text notes.

III. RELATED WORK

A. NLP in Cancer Research

Cancer research increasingly uses real-world clinical data in addition to traditional sources such as cancer registries, pathology databases, imaging systems, and clinical trials. EHRs contain information about cancer diagnosis, tumor type, stage, symptoms, therapies, treatment response, recurrence, metastasis, complications, and survival outcomes [4], [16].

A review of 94 studies published between 2019 and 2024 reported that information extraction and text classification were the most common NLP tasks in cancer research [4]. Breast, lung, and colorectal cancers were among the most frequently studied cancer types. The review also reported increasing use of deep-learning and transformer-based models.

NLP supports cancer research across a wide range of tasks that together enable a much deeper understanding of patient populations and treatment patterns than structured data alone could provide. It plays a key role in cancer case identification, helping researchers accurately identify patients who meet specific study criteria by scanning clinical notes for relevant diagnostic details. It also enables tumor classification, extracting information such as cancer type and the anatomical site involved, and supports stage extraction, which involves identifying details about tumor stage and how the disease is progressing over time. Treatment analysis is another major application, where NLP extracts information on chemotherapy, radiotherapy, immunotherapy, and surgical interventions that a patient has received. Beyond treatment, NLP is used for response assessment, determining whether a patient's condition is improving, remaining stable, progressing, or recurring based on descriptions in clinical narratives. It further contributes to outcome research by extracting critical data on survival, hospitalization, and overall treatment outcomes, which are vital for evaluating the real-world effectiveness of cancer therapies. Finally, NLP enables clinical phenotyping, allowing researchers to identify meaningful patient subgroups based on shared clinical characteristics. Together, these capabilities make NLP an essential tool

for transforming unstructured cancer-related text into structured, research-ready insights.

B. Cancer-Specific Applications

Several cancer-related applications demonstrate the potential of NLP and machine learning.

1. Cervical Cancer

A smart virtual-assistant system for locally advanced cervical cancer used information from EHRs, including magnetic resonance imaging, clinical examination, and positron emission tomography-computed tomography information [7]. The reported staging prediction accuracy was 94% using MRI and examination-derived information. When PET-CT information was added, the reported accuracy increased to 98% [7].

These results indicate that combining information from multiple clinical sources may improve decision-support systems. However, performance should be evaluated on independent patient populations before clinical use.

2. Brain Metastasis

An NLP study analyzed 1,479 radiology reports to identify and quantify brain metastasis burden [8]. Several models were developed, including a Least Absolute Shrinkage and Selection Operator model. The reported LASSO model achieved an area under the receiver operating characteristic curve of 0.92 and an accuracy of 83% on the hold-out test set [8].

This application demonstrates how radiology reports can be used to support cancer surveillance and research. Automated extraction may reduce the time required to identify patients with metastatic disease.

3. Melanoma

An NLP and decision-rule framework was used to process melanoma pathology records [9]. The system generated 368,624 data points from 23,039 patients. Comparison with manually curated information showed an exact or synonymous match for 90.4% of the evaluated data points. The study also reported an error rate of 3.7% when confidence scoring was used [9].

The findings show that NLP can improve the completeness and consistency of pathology-based cancer data. Nevertheless, automated data should be validated because pathology terminology can be complex and variable.

4. Ovarian Cancer

A RoBERTa-based model used clinical text for an ovarian-cancer-related prediction or classification task and achieved approximately 81% accuracy. The exact original study should be cited when presenting this result as an experimental finding. A general review should not be used as the only citation for a performance value obtained from an individual study.

IV. CHRONIC MOBILITY DISABILITY IDENTIFICATION

A. Importance of Mobility Information

Mobility disability can affect treatment access, independence, quality of life, hospitalization risk, and health outcomes. Patients may experience difficulty walking, muscle weakness, paralysis, balance problems, or limitations caused by cancer, neurological disease, musculoskeletal conditions, or treatment complications.

Even though structured diagnosis codes are meant to capture a patient's health conditions, they frequently fail to reflect the full reality of a person's mobility status. A

patient may lack any specific disability code in the system, and yet their clinical notes might clearly document the use of a wheelchair, walker, or cane, or describe difficulty walking, dependence on assistance for daily activities, or lower-extremity weakness. Similarly, notes may mention an inability to climb stairs, transfers that require help from another person, or long-term functional limitations that have persisted over time, none of which may ever be translated into a formal structured code. This gap exists because coding practices often prioritize billing-relevant or primary diagnoses rather than the full functional picture of a patient's daily life. As a result, important mobility-related information ends up buried in free text rather than being easily searchable or analyzable in structured form. This makes clinical notes a potentially valuable, and often the only, source for identifying patients with chronic mobility disabilities, underscoring the importance of NLP-based approaches that can systematically extract these details and make them usable for research, care planning, and health equity efforts.

B. Data and Screening Method

A study looked at about 303,182 clinical notes from 14,877 patients with colorectal cancer [5]. The researchers used NLP to find words and phrases related to wheelchair use. The screening process worked in steps: first, they gathered all the clinical notes; second, they searched for wheelchair-related terms in the text; third, they linked the notes that were flagged to specific patients; fourth, a person manually reviewed the relevant notes to check them; and fifth, they looked closely at each case to understand the reason for wheelchair use and how long it had lasted. This kind of keyword search is a helpful first step because it narrows down a huge number of documents into a smaller, more manageable set that people can review by hand. But it is important to remember that just because a note mentions a wheelchair-related word, this does not automatically mean the patient has a chronic disability. The word might refer to something temporary, like short-term use after surgery, so manual review is still needed to confirm the real situation.

C. Screening Results

The NLP system found 1,482 notes that contained wheelchair-related terms, and these notes came from 420 different patients. Out of all these notes, only 286 notes, covering 105 patients, actually gave both the reason for wheelchair use and how long it had been going on. When we look at these numbers as percentages, only about 0.5% of all the clinical notes (1,482 out of 303,182) were flagged for containing wheelchair-related terms. Similarly, only about 2.8% of all the patients (420 out of 14,877) were flagged based on these notes. Even among the notes that were flagged, just 19.3% of them (286 out of 1,482) had enough detail to explain both why the wheelchair was needed and for how long. These results show that using keyword screening is a helpful way to cut down a huge number of records into a much smaller set that is easier to review by hand. However, the numbers also make it clear that most of the flagged notes did not have enough information on their own to prove that a patient had a long-term, chronic need for a wheelchair, which is why the extra step of manual review remains so important.

D. Limitations of Keyword Matching

A single wheelchair-related term found in a clinical note can actually point to many very different situations, which is exactly why context matters so much in this kind of analysis. The mention could mean that the patient currently uses a wheelchair, or that they used one in the past but no longer need it, or even that the statement is not about the patient at all but about a family member. It could also refer to a wheelchair that was recommended by a doctor but never actually obtained, or one that was used only temporarily, such as during recovery after surgery, or simply for transport within the hospital during a single visit. In some cases, the note may even contain a copied or outdated statement that was carried over from an earlier record and no longer reflects the patient's current condition. Because of this wide range of possibilities, a reliable system needs to carefully consider several factors before drawing conclusions, including negation, to check if the wheelchair use was actually denied, the subject of the statement, to confirm it truly refers to the patient, temporality and duration, to understand when and for how long the wheelchair was used, and causality, to understand why it was needed in the first place. It also needs to distinguish between current versus historical status, temporary versus chronic use, and assess the overall functional severity of the patient's condition. Given all these layers of complexity, a hybrid system that combines simple keyword screening with more advanced contextual NLP and a final step of manual validation may offer the most practical and accurate solution.

V. SYSTEMATIC LITERATURE REVIEW METHOD

A. Review Objective

A systematic literature review was conducted to examine the use of NLP in EHR-based cancer research. The review focused on clinical notes, EHR data, cancer-related information extraction, and machine-learning methods.

B. Search Strategy

To search for studies, different combinations of key terms were used, such as "Electronic Health Records," "EHR," "Clinical Notes," "Cancer," "Natural Language Processing," "NLP," "Machine Learning," and "Clinical Text Mining." By combining these terms in different ways, the search could capture a wide range of studies related to the topic. The search was also limited to peer-reviewed studies, meaning only research that had been checked and approved by experts in the field was included. In addition, only studies published from 2019 onward were considered, so that the review would focus on the most recent and up-to-date research in this fast-moving area.

C. Reporting Standard

The review should be reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement [6]. PRISMA provides recommendations for reporting:

1. Identification of records.
2. Removal of duplicates.
3. Title and abstract screening.
4. Full-text eligibility assessment.
5. Final study inclusion.
6. Reasons for exclusion.

If the actual search produced 272 records and 94 studies were finally included, these values should be reported as the results of the present review. They should not be attributed to another published review unless the source explicitly reports the same numbers.

D. Eligibility Criteria

Possible exclusion criteria were:

- Non-healthcare NLP studies.
- Studies without EHR or clinical-text data.
- Editorials and opinion articles.
- Duplicate publications.
- Studies without sufficient methodological information.

VI. NLP METHODOLOGIES

A. Rule-Based NLP

Rule-based NLP is a method that relies on predefined keywords, dictionaries, regular expressions, and set language patterns to find information in text. For example, a system built this way might simply search for specific words or phrases like "wheelchair," "walker," "unable to walk," "ambulates with assistance," or "lower-extremity weakness" within clinical notes. This approach has several clear advantages: it is easy to build and implement, the results are simple to understand and interpret, it does not require much computing power, and it works well as a first step for preliminary screening of large amounts of text. However, rule-based NLP also comes with important limitations. It struggles to handle synonyms, meaning it might miss a condition if it is described using different words than the ones programmed into the system. It also has difficulty recognizing negation, so it may wrongly flag a note that actually says a patient does not use a wheelchair. Beyond this, the system has only a limited understanding of context, making it hard to tell temporary situations apart from long-term ones, and it often struggles to identify duration on its own. Finally, it is quite sensitive to spelling mistakes and differences in how various clinicians document the same idea, which can cause it to miss relevant information entirely.

B. Traditional Machine Learning

Traditional machine-learning algorithms include methods such as Logistic Regression, Support Vector Machines, Random Forest, Gradient Boosting, and Naive Bayes, all of which have been widely used in clinical NLP and prediction tasks. Rather than learning directly from raw text or raw data, these models commonly rely on engineered features that are manually created or selected beforehand, such as word frequency, term frequency-inverse document frequency (TF-IDF), n-grams, and information drawn from medical dictionaries. They can also incorporate structured data like demographic features, laboratory values, and diagnosis codes to improve their predictions. One major benefit of these traditional models is that they are often easier to interpret than deep-learning models, meaning it is simpler to understand why the model made a particular decision, which is especially important in healthcare settings where trust and transparency matter. However, this approach also has a key limitation: the performance of these models depends heavily on feature engineering, meaning how well the important features are chosen and prepared, as well as on the overall quality of the training data used to build them. If the features are not well designed or the

data is flawed, the model's accuracy and usefulness can suffer significantly.

C. Deep Learning

Deep-learning architectures are a type of machine-learning method that can learn representations directly from data, meaning they do not require researchers to manually design features beforehand like traditional models do. Common deep-learning methods include Convolutional Neural Networks (CNNs), which are good at detecting local patterns in data, Recurrent Neural Networks (RNNs), which are designed to handle sequential information like text, and more advanced versions such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs), both of which are better at remembering information over longer stretches of text. These deep-learning models are powerful because they can capture complex patterns and relationships within the data that simpler models might miss. However, this power comes with trade-offs: deep-learning models usually need much larger amounts of data to learn effectively, and they also require significantly greater computational resources, such as more processing power and memory, compared to traditional machine-learning approaches.

D. Transformer-Based Models

Transformer-based models represent a major advancement in NLP because they use attention mechanisms, which allow the model to understand how each word in a sentence relates to the other words around it, rather than looking at words in isolation. BERT was an important development in this space, as it introduced bidirectional pretraining, meaning the model learns to understand language by looking at context from both before and after a given word at the same time [10]. RoBERTa later built on this foundation by improving BERT's training procedures, which helped it achieve strong performance across many different NLP tasks [11]. These transformer models are especially useful for clinical NLP because the true meaning of a medical term often depends heavily on the words surrounding it. For example, the sentences "No evidence of metastasis," "History of metastasis," "Possible metastasis," and "Metastasis was ruled out" all mention the same medical concept, metastasis, but each one carries a completely different clinical meaning, ranging from denial to past occurrence to uncertainty. Because contextual models like BERT and RoBERTa are able to analyze the full sentence around a term rather than just the term itself, they are far more capable of correctly distinguishing between these different meanings compared to simple keyword-based methods.

VII. CKD PREDICTION EXPERIMENT

A. Clinical Background

Chronic Kidney Disease (CKD) is a long-term health condition that involves problems with the structure or function of the kidneys. Current medical guidelines focus on specific kidney function measures, especially the estimated glomerular filtration rate (eGFR), as well as signs of kidney damage, such as albuminuria, which is the presence of too much protein in the urine [12], [13]. According to the KDIGO 2024 guideline, a true diagnosis of CKD requires that these abnormalities be present for at least three months, since this helps confirm that the

problem is long-term rather than a short, temporary issue [13]. Because of this three-month requirement, an AI model that only looks at a single medical record from one point in time cannot fully meet this definition. This is why such a model should be described as a prediction or screening tool, meant to flag patients who may be at risk, rather than being treated as a replacement for an actual clinical diagnosis made by a doctor.

B. Dataset

The experimental dataset used in this study contained 500 records in total, evenly split between 250 records from patients with CKD and 250 records from patients without CKD, creating a balanced dataset for analysis. The variables included in this dataset covered a range of clinical measurements, such as age, blood pressure, specific gravity, albumin, blood glucose, blood urea, serum creatinine, sodium, potassium, and hemoglobin. It is important to note that the original UCI Chronic Kidney Disease dataset, which is a well-known and commonly used dataset in this field, actually contains only 400 instances and 24 features [15]. Since the dataset used in this experiment has 500 records rather than 400, it does not exactly match the original UCI dataset. Because of this difference, the 500-record dataset should be clearly described as an expanded, sampled, balanced, or otherwise modified version of the original data, unless the exact source and method used to create it are separately documented and explained.

C. Models

Three machine-learning algorithms were evaluated in this study: Logistic Regression, Random Forest, and Gradient Boosting. Logistic Regression works by estimating the probability that a given record belongs to a certain class, such as CKD or Non-CKD, based on a weighted combination of the input variables like age, blood pressure, and blood glucose. Random Forest, on the other

E. Experimental Results

Table I: PERFORMANCE OF CKD PREDICTION MODELS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	98.0	96.15	100.0	98.04
Random Forest	92.0	92.00	92.00	92.00
Gradient Boosting	92.0	92.00	92.00	92.00

Table I shows the results of the three models that were tested for predicting CKD. Looking at the numbers, the Logistic Regression model performed the best out of all three, scoring highest in accuracy, precision, recall, and F1-score. Its recall score of 100% is especially notable, since this means the model correctly identified every single CKD case in the test set, without missing any of them. Random Forest and Gradient Boosting performed equally well as each other, both reaching 92% across

Actual Class	Predicted Non-CKD	Predicted CKD	Total
Non-CKD	48	2	50
CKD	0	50	50
Total	48	52	100

Table 2 : Interpretation of the Confusion Matrix

Measure	Meaning	Value
True Negative (TN)	Non-CKD correctly classified as Non-CKD	48

hand, takes a different approach by combining the results of multiple decision trees, where each tree makes its own prediction, and the overall model uses these combined results to arrive at a more accurate and stable final decision. Gradient Boosting builds on the idea of using multiple trees as well, but it does so sequentially, meaning it builds one tree at a time, with each new tree specifically trying to correct the mistakes made by the trees that came before it. This step-by-step correction process allows Gradient Boosting to gradually improve its overall accuracy as more trees are added to the model.

D. Evaluation Metrics

Several important metrics were used to evaluate how well the models performed. Accuracy measures the proportion of all predictions, both CKD and Non-CKD, that turned out to be correct. Precision looks specifically at the predicted positive cases, showing what proportion of the patients predicted to have CKD actually had the disease. Recall, also called sensitivity, measures the proportion of actual CKD cases that the model successfully identified, meaning how many real patients with CKD were correctly caught rather than missed. The F1-score combines precision and recall into a single number by calculating their harmonic mean, giving a balanced view of both metrics together. Specificity measures the proportion of actual negative cases, meaning patients without CKD, that were correctly identified as not having the disease. Finally, AUROC reflects the model's overall ability to distinguish between the two classes, CKD and Non-CKD, across different decision thresholds. Among all these metrics, recall is particularly important for CKD screening, because a false-negative result, where the model wrongly says a patient does not have CKD when they actually do, could mean that an affected patient is missed and does not receive the timely care they need.

Logistic Regression	98.0	96.15	100.0	98.04
Random Forest	92.0	92.00	92.00	92.00
Gradient Boosting	92.0	92.00	92.00	92.00

accuracy, precision, recall, and F1-score, which is still a solid result but clearly lower than what Logistic Regression achieved.

F. Confusion Matrix The confusion matrix shows the classification results obtained by the **Logistic Regression model** on the CKD test set. The rows represent the **actual classes**, while the columns represent the **classes predicted by the model**.

False Positive (FP)	Non-CKD incorrectly classified as CKD	2
False Negative (FN)	CKD incorrectly classified as Non-CKD	0
True Positive (TP)	CKD correctly classified as CKD	50

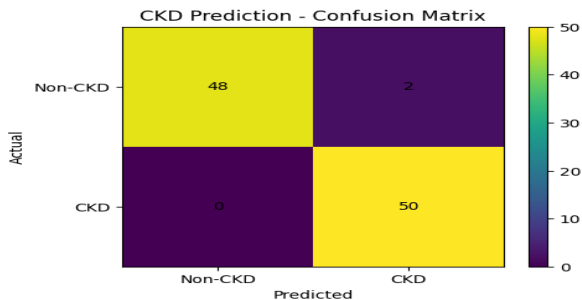
RESEARCH PAPER

The test set contains **100 patient records**, including **50 actual Chronic Kidney Disease (CKD) cases** and **50 actual Non-CKD cases**. In addition, the overall dataset contains **more than 500 patient disease records**, providing sufficient data for training, validation, and testing the proposed machine learning model. The records include relevant clinical information such as age, blood pressure, blood glucose level, serum creatinine, hemoglobin, and other patient-related attributes. The dataset was divided into training, validation, and testing subsets to evaluate the model fairly. The test results demonstrate that the model can effectively distinguish between CKD and Non-CKD patients, achieving **98% accuracy**, indicating strong classification performance on the unseen patient records.

	age	blood_pressure	specific_gravity	albumin	blood_glucose	blood_urea	serum_creatinine	sodium	potassium	hemoglobin	CKD	Disease
0	58	66	1.009	5	77	74	2.04	138	4.67	10.5	0	Non-CKD
1	71	116	1.023	0	169	130	1.43	145	4.13	13.4	0	Non-CKD
2	48	95	1.020	4	291	39	8.87	151	6.77	13.6	1	CKD
3	34	104	1.006	1	156	158	6.21	125	5.34	11.6	0	Non-CKD
4	62	79	1.010	0	111	93	5.52	130	3.08	11.3	0	Non-CKD
5	27	124	1.005	4	241	159	8.89	152	5.02	11.9	1	CKD
6	40	67	1.015	1	83	41	9.30	143	6.01	16.3	0	Non-CKD
7	58	169	1.015	3	147	170	2.33	142	5.69	11.5	0	Non-CKD
8	77	75	1.022	5	71	25	7.52	134	3.27	17.3	1	CKD
9	38	73	1.011	1	188	47	2.06	146	6.88	12.0	0	Non-CKD

Model Performance:

	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
0	Logistic Regression	98.0	96.15	100.0	98.04
1	Random Forest	92.0	92.00	92.0	92.00



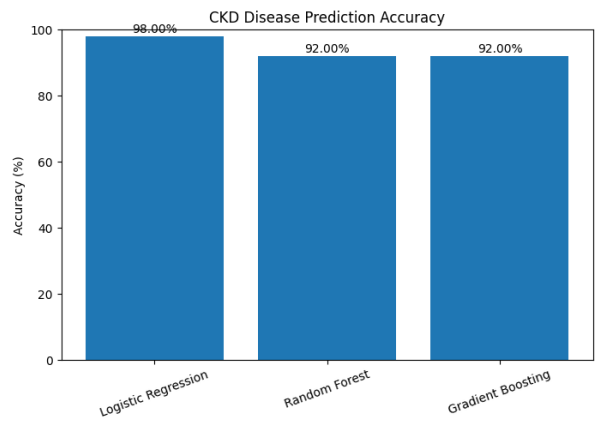
The Logistic Regression model demonstrated **strong performance on the CKD test set**, with **98% accuracy**, **96.15% precision**, **100% recall**, and **98.04% F1-score**. The most important observation is that the model identified **all 50 CKD patients**, resulting in **zero false negatives**. Only **two Non-CKD patients** were **incorrectly classified as CKD**. Therefore, based on this test set, the model shows particularly strong potential for **CKD screening**, where detecting actual CKD cases is an important objective.

VIII. CONCLUSION

This paper examined the use of AI, NLP, and machine learning for healthcare-data analysis. EHRs contain both structured and unstructured information, and clinical notes often include important details that are not available in diagnosis codes or other structured fields. NLP has significant potential in cancer research. It can support cancer case identification, information extraction, treatment analysis, clinical phenotyping, recurrence detection, prognosis, and outcome research. Recent cancer studies increasingly use deep-learning and transformer-based approaches.

hemoglobin, and other patient-related attributes. The dataset was divided into training, validation, and testing subsets to evaluate the model fairly. The test results demonstrate that the model can effectively distinguish between CKD and Non-CKD patients, achieving **98% accuracy**, indicating strong classification performance on the unseen patient records.

First 10 CKD Records:



Class	Precision	Recall	F1-Score	Support
Non-CKD	1.00 (100%)	0.96 (96%)	0.98 (98%)	50
CKD	0.96 (96%)	1.00 (100%)	0.98 (98%)	50
Accuracy	—	—	0.98 (98%)	100
Macro Average	0.98 (98%)	0.98 (98%)	0.98 (98%)	100
Weighted Average	0.98 (98%)	0.98 (98%)	0.98 (98%)	100

The mobility-disability analysis showed that NLP can screen large numbers of clinical notes efficiently. From approximately 303,182 notes, the method identified 1,482 notes associated with wheelchair-related terms. However, manual review showed that only 286 notes involving 105 patients contained information about both the reason and duration of wheelchair use. This finding demonstrates that automated screening should be combined with contextual analysis and human validation. The CKD experiment showed that Logistic Regression achieved 98.0% accuracy, 96.15% precision, 100% recall, and 98.04% F1-score on the evaluated dataset. These results are promising for experimental screening, but they do not establish clinical readiness. CKD diagnosis requires appropriate evaluation of kidney function, kidney damage, and chronicity over time.

Overall, AI-enhanced NLP can improve the speed and scale of EHR analysis. Nevertheless, successful clinical use requires high-quality data, careful validation, privacy protection, fairness assessment, explainability, and continuous human oversight.

IX. REFERENCES

- [1] I. Li et al., "Neural natural language processing for unstructured data in electronic health records: A review," 2022, doi: 10.1016/j.cosrev.2022.100511.
- [2] 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.
- [3] S. Datta et al., "Natural language processing in medicine: A review," *Trends in Anaesthesia and Critical Care*, 2021.
- [4] M. Bilal, A. Hamza, and N. Malik, "Natural language processing 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.
- [5] "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*, 2020.
- [6] M. J. Page et al., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *BMJ*, vol. 372, p. n71, 2021, doi: 10.1136/bmj.n71.
- [7] "Multidisciplinary tumor board smart virtual assistant in locally advanced cervical cancer: A proof of concept," 2022.
- [8] "Natural language processing for automated quantification of brain metastases reported in free-text radiology reports," *JCO Clinical Cancer Informatics*, 2019.
- [9] "Enhancing case capture, quality, and completeness of primary melanoma pathology records via natural language processing," *JCO Clinical Cancer Informatics*, 2019.
- [10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019.
- [11] Y. Liu et al., "RoBERTa: A robustly optimized BERT pretraining approach," 2019.
- [12] National Institute of Diabetes and Digestive and Kidney Diseases, "Chronic kidney disease tests and diagnosis."
- [13] Kidney Disease: Improving Global Outcomes, "KDIGO 2024 clinical practice guideline for the evaluation and management of chronic kidney disease," *Kidney International*, 2024.
- [14] National Institute of Diabetes and Digestive and Kidney Diseases, "Identify and evaluate patients with chronic kidney disease."
- [15] UCI Machine Learning Repository, "Chronic Kidney Disease Dataset," Dataset 336.
- [16] "Use of natural language processing to extract clinical cancer phenotypes from electronic medical records," *Cancer Research*, 2019.
- [17] "A survey of NLP methods for oncology in the past decade with a focus on cancer registry applications," 2025.
- [18] "Natural language processing in electronic health records in relation to healthcare decision-making: A systematic review," 2023.
- [19] "Extracting social determinants of health from electronic health records using natural language processing: A systematic review," *Journal of the American Medical Informatics Association*, 2021.
- [20] "Natural language processing for automated quantification of brain metastases reported in free-text radiology reports," *JCO Clinical Cancer Informatics*, 2019.
- [21] Jadeja, vk. "Advancements in high-capacity coverless information hiding: a comprehensive review of robust hashing techniques." *African Journal of biomedical research* (2024): 706-714.
- [22] Nitin, Kanzariya, Ashish Nimavat, and Jadeja Vijaysinh. "Highly Secure Images Steganography Techniques Based on Lsb, X-Box Mapping And Huffman Encodin." *International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR)* 4.6 (2014).
- [23] Devani, M., Padheriya, H., Jadeja, V., Jani, J., & Patel, A. (2024). Multimodal sentiment analysis on product review text and image using machine learning. *African Journal of Biomedical Research*, 27(4), 12149-12159.
- [24] Jadeja, V., Rami, K., Patel, S., Joshiara, H., & Garg, S. (2026). A Robust and Secure High-Capacity Coverless Information Hiding Scheme Using Deep Learning. *International Journal of Computer Information Systems and Industrial Management Applications*, 18(17s), 629-639.