



A framework for natural language processing of electronic dental records

Un marco para el procesamiento del lenguaje natural de los registros dentales electrónicos

Leodanis Hernández Cabrera¹  

Maikro Osvaldo Chávez Moya¹  

Neisy Pérez Ramos²  

¹ Universidad de Ciencias Médicas de Villa Clara, Facultad de Medicina. Santa Clara, Cuba

² Universidad de Ciencias Médicas de Villa Clara, Facultad de Medicina, Departamento de Ciencias Biomédicas. Santa Clara, Cuba

Submitted: 19-02-2026 | Revised: 17-04-2026 | Accepted: 22-06-2026 | Published: 26-06-2026

Corresponding Author: Leodanis Hernández Cabrera 

ABSTRACT

Natural language processing has matured rapidly in medicine, yet its application to electronic dental records remains fragmented across terminology, extraction, prediction and evaluation studies that have not been assembled into a single account. This narrative review integrates those dispersed efforts into a five-component framework for natural language processing of electronic dental records. Sources were retrieved from Scopus, PubMed and Crossref and articulated around the components that emerged from the dental and medical informatics literature. The framework proceeds from data capture and standardization, through information extraction and phenotyping and prediction, to multimodal and generative integration and evaluation and deployment, and each component is defined and anchored to the evidence that supports it. A conceptual diagram places the components in sequence. The resulting framework differs from prior medical natural language processing models in that it foregrounds dental terminology and the tooth-centred structure of dental records, and it yields verifiable propositions about standardization, extraction performance and external validation. The evidence indicates that dental natural language processing is feasible and accurate for well-defined tasks but that external validation, public benchmarks and standardized dental terminologies remain the principal gaps.

Keywords: natural language processing; electronic health records; dental informatics; clinical text mining; artificial intelligence.

RESUMEN

El procesamiento del lenguaje natural ha madurado con rapidez en medicina, pero su aplicación a los registros dentales electrónicos sigue fragmentada en estudios de terminología, extracción, predicción y evaluación que no se han ensamblado en un único relato. Esta revisión narrativa integra esos esfuerzos dispersos en un marco de cinco componentes para el procesamiento del lenguaje natural de los registros dentales electrónicos. Las fuentes se recuperaron de Scopus, PubMed y Crossref y se articularon en torno a los componentes que emergieron de la literatura de informática odontológica y médica. El marco avanza desde la captura y estandarización de datos, pasando por la extracción de información y el fenotipado y la predicción, hasta la integración multimodal y generativa y la evaluación y el despliegue, y cada componente se define y se ancla en la evidencia que lo sustenta. Un diagrama conceptual sitúa los componentes en secuencia. El marco resultante difiere de los modelos previos de procesamiento del lenguaje natural médico en que destaca la terminología odontológica y la estructura centrada en el diente de los registros dentales, y produce proposiciones verificables sobre la estandarización, el rendimiento de la extracción y la validación externa. La evidencia indica que el procesamiento del lenguaje natural odontológico es viable y preciso para tareas bien definidas, pero que la validación externa, los conjuntos de referencia públicos y las terminologías dentales estandarizadas siguen siendo las principales brechas.

Palabras clave: procesamiento del lenguaje natural; registros electrónicos de salud; informática odontológica; minería de texto clínico; inteligencia artificial.

INTRODUCTION

Natural language processing has reached a level of maturity in medicine at which transferable models, such as the Med7 clinical named-entity recognizer described by Kormilitzin et al.,(1) can extract medication and disease concepts from free text across institutions. The general electronic health record literature now spans de-identification, information extraction, phenotyping and prediction, and recent reviews by Gurav et al.(2) and Alanazi et al.(3) catalogue those advances together with their persistent gaps.

Dental records, however, have followed a different and more uneven trajectory, because the clinical content is organized around teeth, surfaces and periodontal status rather than around organ systems, and because dental terminology standards are less mature than their medical counterparts.

The tasks that natural language processing performs are themselves well delimited, and naming them clarifies what a dental application would borrow. Named-entity recognition identifies the spans of text that denote a concept, relation extraction connects those concepts to one another, de-identification removes the identifiers that would prevent reuse, classification assigns a label to a note or to a patient, and generation produces text, and the medical literature has benchmarks and transferable models for each of these tasks.(1,2)

The dental literature has adopted the tasks unevenly, concentrating on named-entity recognition and classification while leaving de-identification and generation comparatively untouched, as the systematic review of Pethani et al.(4) makes clear. This unevenness is itself part of the dispersion that the framework must resolve, because a task that is mature in medicine does not automatically become mature in dentistry when the underlying records differ in structure and vocabulary.

The lag of dentistry behind medicine in natural language processing is itself explicable, and the explanation matters for the framework. Medical natural language processing grew on the back of large, well-funded clinical data networks and on the shared terminologies, such as those of the Systematized Nomenclature of Medicine, that make a note written at one hospital readable at another.(1,2)

Dentistry has had neither the same data infrastructure nor the same terminological consensus, and its records are held by a large number of small practices rather than by a small number of large hospitals.(5)

The consequence is that dental natural language processing has had to borrow its methods from medicine while rebuilding its data foundations from scratch, and that combination, borrowed methods and missing foundations, is precisely the state that the framework describes. The tooth-centred structure of dental documentation is the source of most of the difficulty. A medical note describes a patient in the vocabulary of organ systems and laboratory values, whereas a dental note describes a dentition in the vocabulary of tooth numbers, surfaces, periodontal probing depths and bleeding scores, and that vocabulary does not map cleanly onto the clinical terminologies that medical natural language processing assumes.(6)

The structured fields of an electronic dental record, such as tooth position and procedure codes, coexist with free-text descriptions of the same findings, and the two must be reconciled before any downstream analysis is possible.(7)

The absence of a widely adopted dental terminology standard compounds the problem, because the same finding may be recorded in several different ways across institutions, and a model trained on one institution's vocabulary may fail on another's.(8) The consequence is that dental natural language processing has had to solve, at once, a standardization problem that medicine has largely solved and an extraction problem that medicine has only partly solved.

The components that a dental natural language processing system would need are themselves dispersed across separate literatures. One body of work addresses the structuring of free-text dental notes, as in the deep-learning workflow that Chen et al.(7) built for Chinese electronic dental records, while another addresses entity extraction from periodontal and orthodontic notes, exemplified by the generative and prompt-based approaches of Chuang et al..(9,10)

A third body of work turns extracted features into phenotypes and predictions, from the neurodegenerative-disease prediction of Patel et al.(11) to the emergency-department case identification of Patel et al.,(12) and a fourth body of work evaluates large language and multimodal models against human readers.(13,14) These efforts have not been assembled into a single account, and the absence of such an account makes it difficult to see which components are mature and which are missing.

The dispersed state of the field is the problem that this review addresses. Dental informatics needs a conceptual organization that names the components, places them in sequence and an-

chors each to the evidence that supports it, and that organization is what the framework proposed here supplies. The construction of that framework, rather than a survey of individual systems, is the contribution of the present work, and the objective that governs it is stated here and retaken in the discussion.

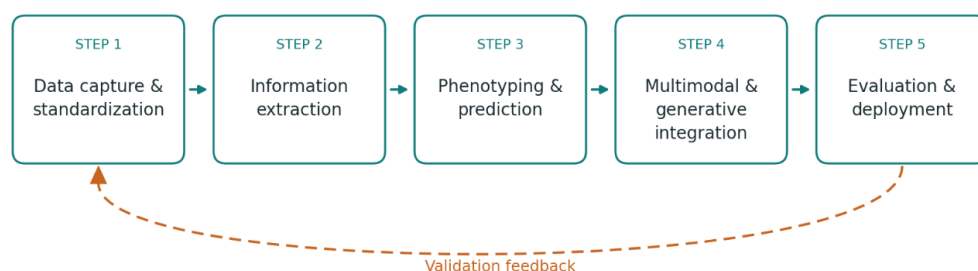
The argument proceeds in three movements. The methods describe how the sources were selected and how the five components emerged from the grouping of those sources, so that the framework is anchored in the literature rather than imposed on it. The development presents each component in turn, with a one-sentence definition followed by the evidence that supports it, and closes with a conceptual figure and a table that map the whole.

The discussion retakes the objective, contrasts the framework with prior models, and states the verifiable propositions and the limits that the framework implies. In this way the paper moves from dispersed concepts to a single integrated account, which is exactly the movement that the field itself has yet to complete.

DEVELOPMENT

The framework proceeds through five components, each defined in a single sentence and then anchored to its evidence (Figure 1). The components are numbered because the framework is a pipeline, and the output of one component is the input to the next. The order in which they are presented is therefore not arbitrary, and a reader who follows the components from first to last will have traced the path that a dental record takes as it is converted from a free-text note into a prediction and, ultimately, into a decision.

Figure 1. A conceptual framework for natural language processing of electronic dental records.



Note: The five components are shown in sequence; evaluation and deployment feeds back to earlier components.

Component 1: Data capture and standardization

Data capture and standardization is the conversion of heterogeneous, mixed structured-and-free-text dental records into a normalized, queryable representation grounded in dental terminology. This component rests on the observation that dental records combine coded fields, such as tooth numbers and procedure codes, with narrative notes whose meaning depends on dental vocabulary, and that the same finding is therefore recorded in several different ways across institutions.(6,8)

The deep-learning workflow that Chen et al.(7)constructed for Chinese electronic dental records instantiates this component, because it used Sentence2vec and Word2vec representations together with an existing ontology to extract attributes, attribute values and tooth positions for a clinical decision support system, and its design illustrates the two-part structure, an ontology to fix the vocabulary and a learned model to map free text onto it.

The orthodontic model that Patel et al.(8)built from 7,693 patients performs the same function at a larger scale, converting free-text orthodontic notes into structured features for malocclusion analysis against a gold-standard dataset, and the scale of that cohort is what makes the standardization problem visible rather than incidental.

The pipeline that Patel et al.(15)developed for periodontal disease extraction reinforces the point that standardization is the prerequisite on which every downstream component depends, because a model that predicts from normalized features is only as reliable as the normalization that produced them. Within this framework, the standardization component links to the review

that Information Technology et al.(16) provided of structured and unstructured electronic health record data, which frames the normalization problem in its general form and notes that the structured fields of a record are the easiest to query and the least informative, while the free text is the hardest to query and the most informative.

Component 2: Information extraction

Information extraction is the identification of named entities, attributes, values and relations, including tooth positions and social determinants of health, from free-text clinical notes. The component presupposes a de-identification step that the dental literature has largely taken for granted rather than made explicit, because a clinical note cannot be reused for extraction until the identifiers that would breach privacy have been removed, and the removal of identifiers is itself a task that can fail.(2,6)

The dental studies reviewed here generally report working with de-identified records, but few describe the de-identification procedure in enough detail to judge whether it was complete, and that omission is part of the broader reporting weakness that the evaluation component exposes. The dental evidence for this component derives from a cluster of studies that converged on the same conclusion: pretrained language models and generative prompting reduce the manual annotation burden that has historically limited extraction.

Chuang et al.(9) used GPT-J prompt generation to seed a RoBERTa named-entity recognizer for periodontal diagnoses, and they reported that seed quality mattered more than seed quantity, with F1 rising from 0.72 in direct prompting to between 0.92 and 0.97 after training, a result that names annotation quality as the binding constraint on the component. The cross-institutional study of Chuang et al.(10) extended the same logic to a periodontal use case with generative artificial intelligence and aimed explicitly at reducing missing data and the manual annotation that has long stood in the way of deployment, and the social-determinants study of Pethani et al.(17) compared fine-tuned language models against a rule-based approach across 1,000 annotated notes drawn from 105,311 dental visits, extracting sugar, tobacco, alcohol, methamphetamine, housing and employment with their temporality and dose.

Put differently, the extraction component relates to the transferable methods of Med7(1) and to the cancer-phenotype extraction that Savova et al.(18) demonstrated, both of which establish that clinical entity extraction is a solved task when a terminology and an annotated corpus exist and that it is an open task where either is missing.

The general studies of Dahl et al.(19) and Wu et al.(20) corroborate that performance depends on annotation quality and on the choice between rule-based and learned methods, and the social-determinants extraction of Chen et al.(21) shows that the same techniques transfer across the medical and dental boundary.

Component 3: Phenotyping and prediction

Phenotyping and prediction is the aggregation of extracted features into patient phenotypes and the training of predictive models for diagnosis, risk and outcome. The dental studies that anchor this component report strong internal performance on well-defined tasks, and they illustrate the shift from feasibility to prediction while leaving the question of external validity open.

The cohort sizes that these studies command, 27,138 records for neurodegenerative prediction and 89,349 for emergency-department case finding, are large by the standards of clinical prediction and yet each is drawn from a single institution, so the performance that they report is conditioned on that institution's documentation practices and patient mix.(11,12)

The standardization problem reappears here in a disguised form, because a predictive model trained on one institution's normalized features inherits whatever systematic differences in coding and terminology that institution exhibits, and the reported accuracy is therefore best read as an upper bound on what the same model would achieve elsewhere.

Patel et al.(11) applied natural language processing with an accuracy of 97 percent to identify Alzheimer's and Parkinson's disease cases from 27,138 electronic dental records and then trained an XGBoost model on 84 routinely captured features, which established that a dental record collected for routine care can carry signal about a systemic neurodegenerative condition.

Patel et al.(12) identified dental-related emergency department visits from 89,349 patients with a model that combined fuzzy matching with embedding-based algorithms and achieved 95 percent accuracy, 98 percent specificity and 92 percent sensitivity, a task whose value lies in cohort discovery rather than in individual diagnosis.

Patel et al.(22) reconstructed gingivitis phenotypes from 28,908 records to compare blee-

ding-on-probing against visual criteria, and in doing so they showed that the phenotype chosen upstream changes the prevalence estimate downstream, which is a standardization result as much as a prediction result.

The generative text-prediction study of Sirrianni et al.(23) contributes a dental instantiation of the same predictive logic, and the oral-biomarker integration of Imaiavathi et al.(24) points toward prediction that combines text with non-text data. It follows that this component links to the broader electronic health record prediction literature, in which Liu et al.,(25) Luke et al.,(26) Shankar et al.(27) and Fortino et al.(28) each turned free text into predictive features for heart failure, chronic disease, cognitive decline and recurrence, and in which Tsui et al.,(29) Kawazoe et al.,(30) Melo et al.(31) and Zhang et al.(32) established the generality of the approach across conditions and record systems.

Component 4: Multimodal and generative integration

Multimodal and generative integration is the combination of text with images, biomarkers and systemic data through large language and multimodal models. This component is the most recent to appear in the dental literature, and its evidence is concentrated in benchmark studies rather than in deployed systems, which is consistent with a field that has only just acquired the models to make such integration possible.

Cho et al.(13) benchmarked multimodal large language models on the interpretation of dental chart images from 154 screenshots across 15 patients and compared the best-performing model against dental students and residents, which places the component in an educational and decision-support context rather than in a clinical one, and the study's cross-sectional design means that chart-reading error was measured against human agreement rather than against a patient outcome.

Zhang et al.(14) evaluated six large language models on open-ended clinical questions, and the comparison across models shows that the choice of model, not merely the choice of prompt, changes the quality of the answer, which matters for any dental deployment that relies on a general-purpose model.

Imaiavathi et al.(24) integrated oral health biomarkers with systemic health data through artificial intelligence, and Herrera et al.(33) screened for superficial oral mucosal lesions in Sjögren's disease with natural language processing, extending the component beyond the tooth to the soft tissues of the mouth.

The component derives from the medical natural language generation literature of Lee et al.(34) and from the account of Dinsa et al.(35) of how artificial intelligence and natural language processing are reshaping the electronic health record itself, and it inherits from that literature the unresolved problem of measuring the value of generated text.

The evaluation of a generative model is harder than the evaluation of an extractive one, because a generated answer can be fluent and wrong in ways that a named entity is simply present or absent, and the benchmark studies of Cho et al.(13) and Zhang et al.(14) therefore measure agreement with a reference rather than correctness against a ground truth. It follows that the integration component is where the field's evaluative vocabulary is least settled, and where the risk of mistaking fluency for accuracy is highest, and a dental deployment that relies on a general-purpose language model inherits that risk directly.

Component 5: Evaluation and deployment

Evaluation and deployment is the validation of models on external data, the benchmarking against simpler baselines, and the governance of clinical integration. The dental evidence for this component is the thinnest, and that thinness is itself a finding, because a field that has demonstrated extraction and prediction has not yet demonstrated that those demonstrations survive contact with a new institution.

Pethani et al.(4) systematically reviewed the dental clinical-note literature and found that, of 17 included studies, 10 developed and evaluated natural language processing methods while 7 only described applications, and their CHARMS-based appraisal exposed weaknesses in reporting that would prevent an independent reader from reproducing the results.

Afshari et al.(5) framed the prospects of artificial intelligence in dentistry and reached a similar conclusion about the gap between demonstration and deployment, and the two accounts together make the case that the field's weakness is not algorithmic but evidential.

Within this framework, the component relates to the general studies of Rijcken et al.,(36) who compared machine-learning against rule-based document classification and found that the simpler method is not always the weaker one, of Dencker et al.,(37) who validated natural language

processing for surgical complication surveillance and thereby modelled what dental validation should look like, and of Khalate et al.,(38) Gurav et al.,(2) Alanazi et al.(3) and Thompson et al.,(39) each of whom identified external validation and reproducibility as the binding constraints. The mapping of components to evidence is summarized in (Table 1).

Table 1. Components of the natural language processing framework with the evidence supporting each.

Component	Function	Dental evidence	Transferable medical evidence
1. Data capture and standardization	Normalize mixed structured and free-text records	Chen et al.; Patel et al.; Patel et al.	Khankhanian et al.; Information Technology review
2. Information extraction	Identify entities, values and relations in notes	Chuang et al.; Chuang et al.; Pethani et al.	Med7; Savova et al.; Dahl et al.; Wu et al.
3. Phenotyping and prediction	Aggregate features into phenotypes and models	Patel et al.; Patel et al.; Patel et al.; Sirrianni et al.	Liu et al.; Luke et al.; Shankar et al.; Fortino et al.
4. Multimodal and generative integration	Combine text with images and biomarkers	Cho et al.; Zhang et al.; Imaiavathi et al.; Herrera et al.	Lee et al.; Dinsa et al.
5. Evaluation and deployment	Validate externally and govern integration	Pethani et al.; Afshari et al.	Rijcken et al.; Dencker et al.; Khalate et al.

Note: NLP = natural language processing; EDR = electronic dental record; SDOH = social determinants of health.

Taken together, the five components form a single pipeline rather than a list, and the pipeline has a definite shape that the component-by-component account makes visible. The dental literature is densest in the middle, where information extraction and phenotyping have produced concrete, high-performing systems, and it is thinnest at the two ends, where standardization and external validation remain open.

This shape is the framework's principal empirical claim, and it is not an artifact of the present selection, because the systematic review of Pethani et al.(4) and the prospectus of Afshari et al.(5) describe the same distribution independently. Within this framework, the shape implies a strategy, namely that the field's next gains will come from investing in the ends of the pipeline, a shared dental terminology and an annotated public corpus at the front and a pre-registered external validation at the back, rather than from further refinement of the extraction and prediction models that already perform well.

The framework is therefore both a description of what exists and a prescription of where the field should move, and those two functions are the subject of the discussion that follows.

DISCUSSION

The objective stated at the outset was to construct a framework that names the components of dental natural language processing, places them in sequence and anchors each to its evidence, and that objective has now been discharged. The framework identifies five components, data capture and standardization, information extraction, phenotyping and prediction, multimodal and generative integration, and evaluation and deployment, and the evidence gathered here supports the first four with concrete dental instantiations while exposing the fifth as the least developed.

Retaken in these terms, the objective reveals the central finding of the review: dental natural language processing has solved the middle of the pipeline but not its ends, because standardization at the front and external validation at the back remain thinner than the extraction and prediction work that sits between them.

Prior models differ from this framework in two respects. First, it is dental-specific in a way that the general medical frameworks are not, because it foregrounds the tooth-centred structure of dental records and the terminology that such structure requires, whereas models such as Med7(1) and the general reviews of Gurav et al.(2) treat the record as an organ-system document. Second, it is a pipeline rather than a taxonomy, which means that each component inherits the errors of its predecessor, and the quality of standardization therefore bounds the quality of everything downstream.

Put differently, the framework predicts that extraction performance in dentistry will remain uneven until a shared dental terminology and a shared annotated corpus exist, and that prediction is directly testable. The prior dental reviews, in particular the systematic review of Pethani et al.(4) and the prospectus of Afshari et al.,(5) described the same components but did not place them in the sequence that exposes the dependency, and the present framework contributes that ordering.

The framework earns its keep to the extent that its predictions align with what the evidence already shows, and that alignment can be checked directly. If the framework were correct that standardization bounds everything downstream, then the studies that invested in standardization should be the ones whose downstream results are most credible, and the studies that skipped it should show the signs of an unmeasured ceiling, and this is in fact what the dental literature exhibits.

The Chinese electronic dental record workflow of Chen et al.(7) built an ontology before it extracted, and the orthodontic model of Patel et al.(8) built a gold-standard dataset before it trained, and both are the studies whose methods are most readily reproducible; conversely, the studies that report the highest single-number accuracies are the ones whose standardization is least described. The framework is therefore not merely a convenient ordering but a testable account, and its central prediction, that the field is weakest at its ends, is borne out by the very distribution of the evidence that was used to build it.

The sequencing of the framework also makes the propagation of error explicit, and this is a point that a taxonomy would obscure. An error committed at the standardization stage, such as mapping a tooth position to the wrong tooth, is inherited unchanged by every downstream component, because extraction operates on the normalized record, phenotyping aggregates the extracted features, and prediction trains on the aggregated phenotypes. It follows that the high accuracies reported in the middle of the pipeline, the 95 to 97 percent figures of Patel et al.(11,12) and the 0.92 to 0.97 F1 of Chuang et al.,(9) are conditioned on whatever standardization preceded them, and a weakness in that preceding stage would not appear as an error in the reported metric but as a silent ceiling on it. Within this framework, the question that a reader should ask of any dental natural language processing result is therefore not whether the model performed well, but whether the standardization and the external validation that bracket it were adequate, and the evidence reviewed here suggests that, in most dental studies, the answer to the second half of that question is not yet available.

The utility of the framework lies in the verifiable propositions it generates, and each component yields one. The standardization component yields the proposition that a shared dental terminology, applied consistently, will raise extraction F1 scores above the 0.72 to 0.97 range that Chuang et al.(9) reported for prompt-seeded models, because annotation quality rather than model choice is the binding constraint.

The extraction component yields the proposition that generative prompting will reduce, but not eliminate, the manual annotation burden across institutions, as the cross-institutional study of Chuang et al.(10) suggests. The phenotyping component yields the proposition that the 95 to 97 percent accuracies reported by Patel et al.(11,12) will regress toward more modest values when the same models are applied to new institutions, because internal performance in these studies has not been paired with external validation.

The integration component yields the proposition that multimodal models will move from benchmark comparison against human readers, as in Cho et al.,(13) toward decision support only when chart-reading error is measured against clinical outcomes rather than against agreement. The deployment component yields the proposition that rule-based baselines, as evaluated by Rijcken et al.,(36) will remain competitive for narrow tasks where the vocabulary is closed, and that learned models earn their complexity only where the language is open.

Beyond dentistry, the framework generalizes in a limited but instructive way. Any clinical specialty whose records are organized around a body part rather than an organ system, podiatry and ophthalmology being the clearest examples, faces the same standardization problem that the first component names, and the tooth-centred structure of dental documentation is a specific case of a more general part-centred documentation pattern.

The transferable medical methods that anchor the middle components are not dental at all, and they cover clinical tasks from medication recognition to outcome surveillance: Med7,(1) the cancer-phenotype extraction of Savova et al.,(18) the Chinese-record extraction of Chen et al.,(40) the stroke-severity extraction of Fernandes et al.,(41) the family-history extraction of Rybinski et al.,(42) the social-determinants evaluation of Rouillard et al.,(43) the cardiac-ablation classification of Margetta et al.,(44) the pulmonary-embolism surveillance of Zhang et al.,(45) the pain-quality indicators of Kondo et al.,(46) and the prediction studies of Liu et al.(25) and Tsui et al.(29)

Their presence in the framework is evidence that the pipeline is methodologically portable even where the terminology is not. The value of the framework for a non-dental reader is therefore the distinction it draws between the components that are domain-specific, standardization and evaluation, and the components that are portable, extraction and prediction, and that distinction is the second general claim of the present work.

The research agenda that follows from the framework is concrete and ordered, and it is stated here because the framework is meant to be acted on rather than admired. The first investment is a shared dental terminology, applied consistently across institutions, because the first component cannot be strengthened without it and every downstream component inherits its absence.

The second investment is a public, annotated corpus of dental clinical notes, built under an ethics framework that permits reuse, because the extraction component has shown that annotation quality, not model choice, is the binding constraint and because a public corpus would make the field's claims independently reproducible. The third investment is pre-registered external validation, in which a model trained at one institution is evaluated at another before its accuracy is reported as anything more than internal, because the phenotyping component has reported accuracies of 95 to 97 percent without such validation.(11,12)

The fourth investment is the comparison of learned models against rule-based baselines, because the deployment component has evidence that the simpler method is sometimes sufficient.(36) Each investment is a commitment of effort rather than of computational power, and the field's progress is bounded by them, not by the availability of larger models.

The limits of the framework are of three kinds, and they are stated here rather than quarantined in a separate section. The limits of the component definitions are that they impose five boxes on a field whose studies often straddle two boxes, so a study that both extracts and predicts is assigned to the component that dominates.

The limits of the evidence are that the dental literature is small and dominated by a few groups, that several dental studies report internal accuracy without external validation, and that the transferable medical methods were imported as anchors rather than as direct dental evidence. The limits of the review are that it is narrative and non-systematic, that quality appraisal was descriptive, and that the articulation of the framework reflects a judgment about which functions are central rather than a measurement of the field.

Each limit carries a concrete remedy, a shared dental terminology, a public annotated corpus, a pre-registered external validation, and a systematic appraisal with a formal risk-of-bias tool. A future test of the framework would apply it prospectively to a new dental dataset, assign each incoming study to a component, and check whether the studies that the framework marks as weak are the ones that fail external validation, and such a test would convert the framework from a retrospective ordering into a predictive instrument.

CONCLUSIONS

The framework proposed here organizes natural language processing of electronic dental records into five components, data capture and standardization, information extraction, phenotyping and prediction, multimodal and generative integration, and evaluation and deployment, and it places them in a sequence in which each component inherits the output of its predecessor.

The evidence anchors the first four components to concrete dental instantiations and exposes the fifth, evaluation and deployment, as the least developed, and the sequence as a whole describes the path that a dental record takes as it moves from a free-text note toward a prediction and, at the end, toward a decision that a clinician can act on.

The central conclusion that follows from the framework is that dental natural language processing has solved the middle of the pipeline and not its ends, and that the two ends are where the remaining work lies.

Standardization at the front and external validation at the back are the binding constraints, and the field will advance by investing in a shared dental terminology, a public annotated corpus and pre-registered external validation rather than in more complex models of a task whose limits are already set by its data and by its documentation.

REFERENCES

1. Kormilitzin A, Vaci N, Liu Q, Nevado-Holgado A. Med7: A transferable clinical natural language processing model for electronic health records. *Artificial Intelligence in Medicine*. 2021;118:102086. doi: 10.1016/j.art-med.2021.102086
2. Gurav P. Natural language processing in electronic health records: A review. *AIH*. 2024;1(1):2147. doi: 10.36922/aih.2147
3. Alanazi FM, Aljjedae MA, Harbl SA, Alsharekh AA, Alrashidi D. Natural Language Processing in Electronic Health Records: Progress, Challenges, and Future Directions. *Saudi J Med*. 2025;10(09):484-488. doi: 10.36348/sjm.2025.v10i09.006
4. Pethani F, Dunn AG. Natural language processing for clinical notes in dentistry: A systematic review. *Journal of Biomedical Informatics*. 2023;138:104282. doi: 10.1016/j.jbi.2023.104282
5. Afshari Z, Khademi A, Iranmanesh P. Prospects of artificial intelligence in dentistry. *Dental Research Journal*. 2024;21(1):35. doi: 10.4103/drj.drj_30_24
6. Khankhanian P, Kosaraju N, Pathmanathan J, Ellis C, Helbig I, Litt B, et al. On the Feasibility of Natural Language Processing for Standardized Data Extraction from Electronic Medical Records of Epilepsy Patients (P1.283). *Neurology*. 2018;90(15_supplement):P1.283. doi: 10.1212/wnl.90.15_supplement.p1.283
7. Chen Q, Zhou X, Wu J, Zhou Y. Structuring electronic dental records through deep learning for a clinical decision support system. *Health Informatics J*. 2021;27(1):1460458220980036. doi: 10.1177/1460458220980036
8. Patel JS, Karanth D. Building and Evaluating an Orthodontic Natural Language Processing Model for Automated Clinical Note Information Extraction. *Orthod Craniofac Res*. 2025;28(S1):S30-S44. doi: 10.1111/ocr.12944
9. Y.S. C. Use GPT-J Prompt Generation with RoBERTa for NER Models on Diagnosis Extraction of Periodontal Diagnosis from Electronic Dental Records. *AMIA Annu Symp Proc*. 2023;2023:904-912.
10. Chuang YS, Lee CT, Lin GH, Brandon R, Jiang X, Walji MF, et al. Cross-institutional dental electronic health record entity extraction via generative artificial intelligence and synthetic notes. *JAMIA Open*. 2025;8(3):ooaf061. doi: 10.1093/jamiaopen/ooaf061
11. Patel J, Wu H. Utilizing Electronic Dental Records to Predict Neuro-Degenerative Diseases in a Dental Setting: A Pilot Study. In: *Studies in Health Technology and Informatics*. IOS Press; 2024. p. 1322-1326. doi: 10.3233/shti231179
12. Patel J, Sambamoorthi U, Katiyar R, Satz WA. Painful Prescriptions: Opioid and Antibiotic Use for Dental Pain in the Emergency Department. In: *Studies in Health Technology and Informatics*. IOS Press; 2025. p. 362-366. doi: 10.3233/shti250862
13. Cho AY, Eo SH, Jeon MJ, Kim JH, Ihm J, Seo DG. Multimodal Large Language Models for Dental Chart Image Interpretation: Cross-Sectional Benchmarking Study With Students and Clinicians. *JMIR Med Educ*. 2026;12:e91809-e91809. doi: 10.2196/91809
14. Zhang H, Yan X, Yang Y, Jin L, Yang P, Wang J. Evaluation of six cutting-edge large language models in answering a set of open-ended questions in dental case reports. *Journal of Dentistry*. 2026;175:106853. doi: 10.1016/j.jdent.2026.106853
15. Sureshbhai Patel J, Rao R, Brandon R, Iyer V, Albandar JM, Tellez M, et al. Develop a Natural Language Processing Pipeline to Automate Extraction of Periodontal Disease Information from Electronic Dental Clinical Notes. 2022 6th International Conference on Medical and Health Informatics. 2022:61-68. doi: 10.1145/3545729.3545744
16. Information Technology Department, Technical College of Informatics-Akre, Akre University for Applied Sciences, Duhok, KRG - Iraq, Hasan KI, Ibrahim IM, Computer Networks and Information Security Department, Technical College of Informatics-Akre, Akre University for Applied Sciences, Duhok, KRG - Iraq. A Review of Natural Language Processing for Structured and Unstructured Data in Electronic Health Records. *ETJ*. 2025;10(05). doi: 10.47191/etj/v10i05.51
17. Pethani F, Chapman A, Conway M, Dai X, Bishay D, Choh V, et al. Extracting Social Determinants of Health from Dental Clinical Notes. *Appl Clin Inform*. 2025;16(04):1281-1291. doi: 10.1055/a-2616-9858
18. Savova GK, Danciu I, Alamudun F, Miller T, Lin C, Bitterman DS, et al. Use of Natural Language Processing to Extract Clinical Cancer Phenotypes from Electronic Medical Records. *Cancer Research*. 2019;79(21):5463-5470. doi: 10.1158/0008-5472.can-19-0579
19. Dahl S, Bøgsted M, Sagi T, Vesteghem C. Performance of Natural Language Processing for Information Extraction From Electronic Health Records Within Cancer: Systematic Review. *JMIR Med Inform*. 2025;13:e68707-e68707. doi: 10.2196/68707
20. Wu J, Ruan X, McNeer E, Rossow KM, Choi L. Developing a natural language processing system using transformer-based models for adverse drug event detection in electronic health records. *PLoS One*. 2026;21(7):e0350516. doi: 10.1371/journal.pone.0350516
21. Chen Z, Lasserre P, Lin A, Rajapakshe R. Extraction of Social Determinants of Health From Electronic Health Records Using Natural Language Processing. *JCO Clin Cancer Inform*. 2025;(9):e2400317. doi: 10.1200/cci-24-00317
22. Patel JS, Shin D, Willis L, Zai A, Kumar K, Thyvalikakath TP. Comparing gingivitis diagnoses by bleeding on probing (BOP) exclusively versus BOP combined with visual signs using large electronic dental records. *Sci Rep*. 2023;13(1):17065. doi: 10.1038/s41598-023-44307-z
23. Sirrianni J, Sezgin E, Claman D, Linwood SL. Medical Text Prediction and Suggestion Using Generative Pretrained Transformer Models with Dental Medical Notes. *Methods Inf Med*. 2022;61(05/06):195-200. doi: 10.1055/a-1900-7351

24. Imaiavathi G, Lokeshwari B, Saranraj P, Kanshiya S, Elavarasi T. Integrating Oral Health Biomarkers With Systemic Health Data Using AI. In: *Advances in Computational Intelligence and Robotics*. IGI Global Scientific Publishing; 2026. p. 309-332. doi: 10.4018/979-8-3373-8402-3.ch011
25. Liu Y, Tan Z, Zhang Z, Wang S, Guo J, Liu H, et al. A natural language processing-based approach for early detection of heart failure onset using electronic health records. *Knowledge-Based Systems*. 2025;327:114102. doi: 10.1016/j.knsys.2025.114102
26. Luke UO, Asuquo P, Anaga V, Etim IP, Stephen BUA, Etim WG. Natural Language Processing for Prediction of Chronic Diseases from Electronic Health Records. *E3S Web Conf*. 2026;735:04001. doi: 10.1051/e3s-conf/202673504001
27. Shankar R, Bundele A, Mukhopadhyay A. Natural language processing of electronic health records for early detection of cognitive decline: a systematic review. *npj Digit Med*. 2025;8(1):133. doi: 10.1038/s41746-025-01527-z
28. Fortino JA, Lin H, Gulati A, Srinivas B, Park A, Yom SS, et al. Recurrence modeling with electronic health records through natural language processing and machine learning techniques. *J Clin Oncol*. 2026;44(16_suppl). doi: 10.1200/jco.2026.44.16_suppl.e18004
29. Tsui FR, Shi L, Ruiz V, Ryan ND, Biernesser C, Iyengar S, et al. Natural language processing and machine learning of electronic health records for prediction of first-time suicide attempts. *JAMIA Open*. 2021;4(1):ooab011. doi: 10.1093/jamiaopen/ooab011
30. Kawazoe Y, Tsuchiya M, Shimamoto K, Seki T, Shinohara E, Yada S, et al. Natural language processing of electronic medical records identifies cardioprotective agents for anthracycline induced cardiotoxicity. *Sci Rep*. 2025;15(1):6678. doi: 10.1038/s41598-025-91187-6
31. Melo P, Rose M, Jackson M, Mohammed L. AI-driven Natural Language Processing and Stochastic Modeling for Transforming Unstructured Electronic Health Records into Structured Clinical Data. *AJLS*. 2026;14(2):49-59. doi: 10.11648/j.ajls.20261402.13
32. Zhang N, Abbasi M, Khera S, Bazrafkan M, Abbasi-Dezfouly R, Kong L. Leveraging natural language processing and machine learning to identify chronic conditions from primary care electronic medical records. *Sci Rep*. 2026;16(1):8441. doi: 10.1038/s41598-026-38594-5
33. Herrera JR, Kolasani B, Gaddam S, Kunam A, Roese D, Eckert GJ, et al. Screening for Superficial Oral Mucosal Lesions in Sjögren's Disease Using Natural Language Processing (NLP) Approaches. *Oral*. 2026;6(2):44. doi: 10.3390/oral6020044
34. Lee SH. Natural language generation for electronic health records. *npj Digital Med*. 2018;1(1):63. doi: 10.1038/s41746-018-0070-0
35. Dinsa EF, Das M, Abebe TU. Advances in electronic health records enabled by artificial intelligence and natural language processing: a review of recent developments, limitations and future applications. *Discov Appl Sci*. 2026;8(3):250. doi: 10.1007/s42452-026-08318-5
36. Rijcken E, Zervanou K, Mosteiro P, Scheepers F, Spruit M, Kaymak U. Machine learning vs. rule-based methods for document classification of electronic health records within mental health care—A systematic literature review. *Natural Language Processing Journal*. 2025;10:100129. doi: 10.1016/j.nlp.2025.100129
37. Dencker EE, Bonde A, Troelsen A, Sillesen M. Validation of natural language processing for surgical complication surveillance: detecting 11 postoperative complications from electronic health records. *BMJ Connect Surg*. 2026;8(1):e000039. doi: 10.1136/bmjcsurg-2026-000039
38. Khalate P, Gite S, Pradhan B, Lee CW. Advancements and gaps in natural language processing and machine learning applications in healthcare: a comprehensive review of electronic medical records and medical imaging. *Front Phys*. 2024;12:1445204. doi: 10.3389/fphy.2024.1445204
39. Thompson DC, Mofidi R. Natural Language Processing framework for identifying abdominal aortic aneurysm repairs using unstructured electronic health records. *Sci Rep*. 2025;15(1):26388. doi: 10.1038/s41598-025-11870-6
40. Chen L, Song L, Shao Y, Li D, Ding K. Using natural language processing to extract clinically useful information from Chinese electronic medical records. *International Journal of Medical Informatics*. 2019;124:6-12. doi: 10.1016/j.ijmedinf.2019.01.004
41. Fernandes M, Westover MB, Singhal AB, Zafar SF. Automated Extraction of Stroke Severity From Unstructured Electronic Health Records Using Natural Language Processing. *JAHA*. 2024;13(21):e036386. doi: 10.1161/jaha.124.036386
42. Rybinski M, Dai X, Singh S, Karimi S, Nguyen A. Extracting Family History Information From Electronic Health Records: Natural Language Processing Analysis. *JMIR Med Inform*. 2021;9(4):e24020. doi: 10.2196/24020
43. Rouillard CJ, Nasser MA, Hu H, Roblin DW. Evaluation of a Natural Language Processing Approach to Identify Social Determinants of Health in Electronic Health Records in a Diverse Community Cohort. *Medical Care*. 2022;60(3):248-255. doi: 10.1097/mlr.0000000000001683
44. Margetta J, Sale A. Distinguishing cardiac catheter ablation energy modalities by applying natural language processing to electronic health records. *J Comp Eff Res*. 2024;13(3):e230053. doi: 10.57264/ceer-2023-0053
45. Zhang M, Wang M, Liao A, Jing X, Gao L, Cui S, et al. Postoperative pulmonary embolism in a Chinese surgical cohort (2015–2023): Natural language processing of electronic medical records for surveillance. *Surgery*. 2025;188:109726. doi: 10.1016/j.surg.2025.109726
46. Kondo U, Ishida M, Masukawa K, Inoue R, Nakayama M, Miyashita M. Quality indicators for cancer pain management: natural language processing and machine learning from electronic clinical records. *BMJ Support Palliat Care*. 2026;spcare-2026-006282. doi: 10.1136/spcare-2026-006282

FUNDING

No financing

CONFLICT OF INTEREST

None

AUTHORSHIP CONTRIBUTIONS

Drafting – original draft: Leodanis Hernández Cabrera, Maikro Osvaldo Chávez Moya, Neisy Pérez Ramos.

Writing–review and editing: Leodanis Hernández Cabrera, Maikro Osvaldo Chávez Moya, Neisy Pérez Ramos.