ORIGINAL RESEARCH **Open access**

Natural Language Processing Model for Detecting Inconsistencies between Clinical Notes, Problem Lists, Medication Orders, and Billing Codes in Electronic Health Records

Bruno Martins[1*], Lucas Pereira[1], Renata Azevedo[2], Pedro Costa[1]

Abstract

Clinical notes, problem lists, medication orders, and billing codes are core components of the electronic health record. When these components conflict, the record may become less reliable for care delivery, quality measurement, and reimbursement. Current inconsistency detection is largely manual, episodic, and dependent on documentation audits. This approach is difficult to scale across encounters, specialties, and longitudinal records. This article proposes a deep learning NLP model for detecting contradictions among clinical notes, problem lists, medication orders, and billing codes. The goal is to support continuous documentation integrity surveillance. The proposed model uses transformer-based encoders for clinical text and embedding layers for structured coded fields. Cross-attention mechanisms align concepts across EHR modules before classifying consistency relationships. Conceptually, the model could surface discrepancies such as a diagnosis documented in a note but absent from the problem list, or a billing code unsupported by physician documentation. Its output would include an inconsistency category and an interpretable explanation for clinician review. A unified NLP model for cross-module inconsistency detection could improve EHR trustworthiness, documentation quality, and clinical audit workflows. Such a system should be evaluated prospectively before operational deployment.

Keywords Deep learning, Electronic health records, Clinical NLP, Documentation integrity, Inconsistency detection, Problem lists

*Correspondence:
Bruno Martins
bruno.martins@gmail.com
[1] Department of Health Informatics and AI Systems, Faculty of Medicine, University of Minho, Braga, Portugal
[2] Department of Clinical Digital Analytics, Faculty of Engineering, University of Porto, Porto, Portugal

Introduction

Electronic health records are composed of multiple documentation modules that are often updated by different professionals, at different times, and for different purposes. Clinical notes may describe diagnostic reasoning captured in narrative form [1, 2], problem lists may summarize active or historical conditions within structured EHR representations [3, 4], medication orders may imply treated diseases or safety-relevant therapeutic decisions [5, 6], and billing codes may encode diagnoses or procedures for administrative use [7, 8]. These parallel workflows create opportunities for semantic divergence, including unsupported codes, missing active problems, medication–diagnosis mismatches, and outdated documentation. Such inconsistencies can undermine clinical communication, medication safety, quality reporting, and reimbursement integrity [9-11].

Manual review remains the dominant approach for detecting documentation inconsistencies, but chart audits are time-consuming and difficult to sustain across large EHR repositories. Clinical NLP systems can extract diagnoses, medications, assertions, and semantic relationships from free-text documentation [1, 12], while contextual embeddings can strengthen recognition of clinical concepts in note text [2]. Medication extraction and adverse event detection methods further show that clinically meaningful facts can be recovered from narrative documentation [5, 6, 9]. These advances create an opportunity to frame cross-module reconciliation as a scalable model-based task rather than a purely manual documentation review activity [13-15].

Prior research has often addressed isolated components of the broader documentation integrity problem. Automated coding models have linked clinical notes to diagnosis codes [7, 8, 16], medication extraction systems have identified drug concepts in clinical narratives [6, 10], and semantic search systems have mapped clinical concepts from notes into structured representations [17]. Disease-specific NLP pipelines can identify conditions that may be absent from structured fields [18, 19], but these approaches usually focus on one source or one mapping task. A unified model remains needed to jointly reason over whether clinical notes, problem lists, medication orders, and billing codes mutually support, contradict, or omit clinically relevant information [15, 20].

This article proposes a deep learning NLP architecture that jointly analyzes free-text notes, structured problem-list entries, medication orders, and billing codes to detect semantic inconsistencies. Transformer-based biomedical and clinical language models provide a foundation for contextual note representation, while transformer architectures for EHR data suggest how longitudinal and structured clinical information could be encoded together. The proposed model would normalize concepts through clinical vocabularies, align evidence across sources, and classify cross-module relationships as consistent, inconsistent, uncertain, or reviewable. Its purpose is not to replace clinician judgment but to prioritize records requiring review and correction in documentation integrity workflows.

Background

The EHR as a multi-modular documentation system

The EHR functions as a multi-modular documentation environment rather than a single coherent narrative. Clinical notes preserve reasoning and context [1, 14], problem lists summarize diagnostic state within structured EHR data [3], medication orders operationalize treatment decisions [5, 6], and billing codes translate care into administrative categories [7, 8]. Inconsistency can arise when these modules are updated asynchronously, copied forward without reconciliation, or optimized for different clinical and administrative purposes. Large-scale EHR learning studies show the value of integrated records [3, 11], but they also underscore why semantic alignment across modules is essential for dependable clinical and secondary use [4, 15].

NLP for clinical information extraction

Clinical NLP has advanced from rule-based extraction toward hybrid and neural systems capable of identifying concepts, assertions, relationships, and contextual modifiers in free text. CLAMP demonstrates how customizable pipelines can support clinical concept extraction [1], while contextual embeddings improve recognition of entities in narrative documentation [2]. Relation extraction methods can link diseases, findings, medications, and other clinical facts [12], and negation or uncertainty detection is essential for distinguishing present conditions from ruled-out conditions [21]. These capabilities provide the extraction layer required before a model can compare notes with structured problem lists, medication orders, and billing records [14, 15].

Inconsistency and contradiction detection in text

Detecting inconsistency requires more than identifying clinical entities; it requires determining whether one representation entails, contradicts, duplicates, or omits information found elsewhere. Biomedical language models such as BioBERT

support contextual representation of medical text [22], and ClinicalBERT-style embeddings extend this approach to clinical documentation [23]. Transformer-based EHR models further suggest that multi-source clinical information can be represented jointly rather than as disconnected fields [24]. In this setting, inconsistency detection would resemble clinical natural language inference, where note evidence is compared with structured claims embedded in problem lists, medication orders, and billing codes [15, 25].

Clinical concept mapping and ontology alignment

Concept mapping is central to cross-module inconsistency detection because free-text mentions and structured fields often use different surface forms for the same clinical idea. A note may describe “congestive heart failure,” a problem list may contain a mapped diagnosis concept, a medication order may imply treatment for fluid overload, and a billing code may represent a related administrative category. Semantic search systems show how normalized clinical concepts can be surfaced from free text [17], while medication extraction systems demonstrate the value of mapping narrative drug mentions to structured medication representations [5, 6]. Ontology alignment through UMLS, SNOMED CT, RxNorm, and ICD-style representations would therefore allow the model to compare concepts rather than strings [1, 7, 16].

Prior efforts in automated documentation integrity

Automated documentation integrity has been approached through coding prediction, adverse event detection, medication extraction, and disease identification from clinical notes. Explainable coding models can associate note evidence with predicted codes [8], broader code-assignment approaches can model links between clinical text and billing categories [7, 16], and adverse event extraction systems show how clinical narratives can support safety-relevant detection tasks [9, 10]. Disease-specific NLP models can identify conditions that may be missing from structured fields [18, 19], while consumer medication question work illustrates the broader challenge of connecting medication-related text with trusted clinical knowledge [20]. A documentation integrity model would extend these prior efforts by explicitly classifying whether multiple EHR modules are mutually coherent for the same patient encounter.

Model Development Overview

High-level inconsistency detection pipeline

The proposed pipeline begins by extracting clinical concepts from notes, problem lists, medication orders, and billing records, then normalizing these concepts into shared semantic representations. Clinical NLP toolkits and contextual concept extraction methods could support the note-processing stage [1, 2], while semantic search systems could help link extracted mentions to normalized clinical concepts [17]. Each encounter would then be represented as linked evidence units, including note sentences, active problems, medication concepts, and billing codes. Pairwise or multi-document classifiers could determine whether these evidence units are consistent, inconsistent, duplicative, or insufficiently supported [3, 15].

Core input types and alignment strategy

The model would treat clinical notes as token sequences, problem-list items as diagnosis concepts, medication orders as drug concepts, and billing codes as structured administrative labels. Alignment would rely on patient identifiers, encounter dates, section context, and concept mappings that connect free-text mentions to standardized vocabularies [4, 17]. For example, a note sentence documenting diabetes treatment could be aligned with a problem-list diagnosis, an antidiabetic medication order, and a diagnosis code when all refer to the same clinical episode. This alignment strategy reflects prior work on medication extraction [5, 6], automated coding [7, 8, 16], and clinical concept normalization [1, 2].

Design principles

The model should be scalable across hospital services, explainable in clinical terms, and adaptable to local documentation practices. Because clinical workflows are sensitive to alert burden, the system should prioritize discrepancies with plausible clinical or billing relevance rather than flagging every mismatch [13, 25]. It should distinguish documentation errors from clinically appropriate differences, such as historical diagnoses that remain in notes but are no longer active problems [14, 15]. These design principles are consistent with implementation-aware clinical NLP development and with EHR modeling work that treats clinical data as complex, contextual, and workflow-dependent [3, 11, 24].

Data Sources and Feature Engineering

Extraction and standardisation from four EHR modules

The data layer would retrieve progress notes, discharge summaries, problem-list entries, medication orders, and billing codes from EHR data warehouses. Notes would be segmented into clinically meaningful units [14], while structured fields would be standardized by code system, encounter, and timestamp [4]. Medication orders could be normalized to RxNorm-style concepts using medication extraction principles [5, 6], and billing codes could be grouped into diagnosis or procedure categories before comparison with note evidence [7, 16]. Prior work on large-scale EHR modeling shows why such standardization is necessary before downstream deep learning systems can reason across heterogeneous record components [3, 11].

Concept extraction and UMLS normalisation

Clinical NLP pipelines would extract diagnoses, symptoms, medications, procedures, negations, uncertainty markers, and contextual qualifiers from free-text notes. Extracted concepts would be normalized to controlled vocabularies so that lexical variants in notes could be compared with coded problem-list and billing entries [1, 17]. Contextual embeddings and clinical BERT-style representations could improve recognition of semantically equivalent phrases [2, 22, 23], while assertion, negation, and relation features would help distinguish present diagnoses from ruled-out or historical mentions [12, 21]. This feature layer is essential because inconsistency detection depends on whether the model understands both the concept and its clinical status [14, 15].

Building document pairs and inconsistency labels

Training examples would be constructed as document pairs or document groups linking note evidence with problem-list items, medication orders, and billing codes from the same encounter. Expert chart review would define inconsistency categories such as missing diagnosis, unsupported billing code, medication without documented indication, duplicate documentation, or contradiction between note assertion and structured field. Condition-identification and adverse-event NLP studies illustrate how expert-reviewed clinical text can support supervised modeling of clinically meaningful labels [9, 10, 18, 19]. Automated coding studies further show that note–code relationships can be modeled as evidence links, although inconsistency labels would require clinical judgment rather than simple code prediction [7, 8, 16].

A structured typology of cross-module inconsistency categories and their semantic interpretations is provided in **Table 1** to clarify how different discrepancy types are operationalized within the proposed model.

Table 1. Cross-Module Inconsistency Typology and Semantic Interpretation Framework

Inconsistency Type	Source Modules Involved	Semantic Definition	Example Scenario	Clinical Interpretation	Model Output Label

Omission	Notes vs Problem List	Clinically relevant concept present in narrative but absent in structured field	Diabetes described in note but missing from problem list	Documentation incompleteness	Inconsistent
Unsupported Code	Notes vs Billing Codes	Billing code lacks supporting clinical evidence in documentation	CHF billing code without mention in note	Potential coding error	Inconsistent
Medication–Diagnosis Mismatch	Medication Orders vs Notes/Problem List	Medication implies condition not documented	Insulin ordered without diabetes diagnosis	Safety or documentation gap	Inconsistent
Contradiction	Notes vs Structured Fields	Explicit disagreement between sources	Note states “no infection” but infection code present	Logical inconsistency	Inconsistent
Duplication	Across modules or within module	Same concept redundantly recorded	Repeated diagnoses across fields	Redundant documentation	Duplicate
Temporal Misalignment	Across time-linked modules	Concept valid but not aligned temporally	Medication added before diagnosis documented	Sequence inconsistency	Uncertain
Clinically Justifiable Difference	Any	Apparent mismatch explained by clinical context	Historical diagnosis in note but not active	Appropriate variation	Review Required

NLP architecture for inconsistency detection

The proposed hierarchical architecture for cross-module inconsistency detection is illustrated in **Figure 1,** highlighting the sequential flow from multi-source EHR inputs through concept normalization, cross-attention alignment, and explainable classification outputs.

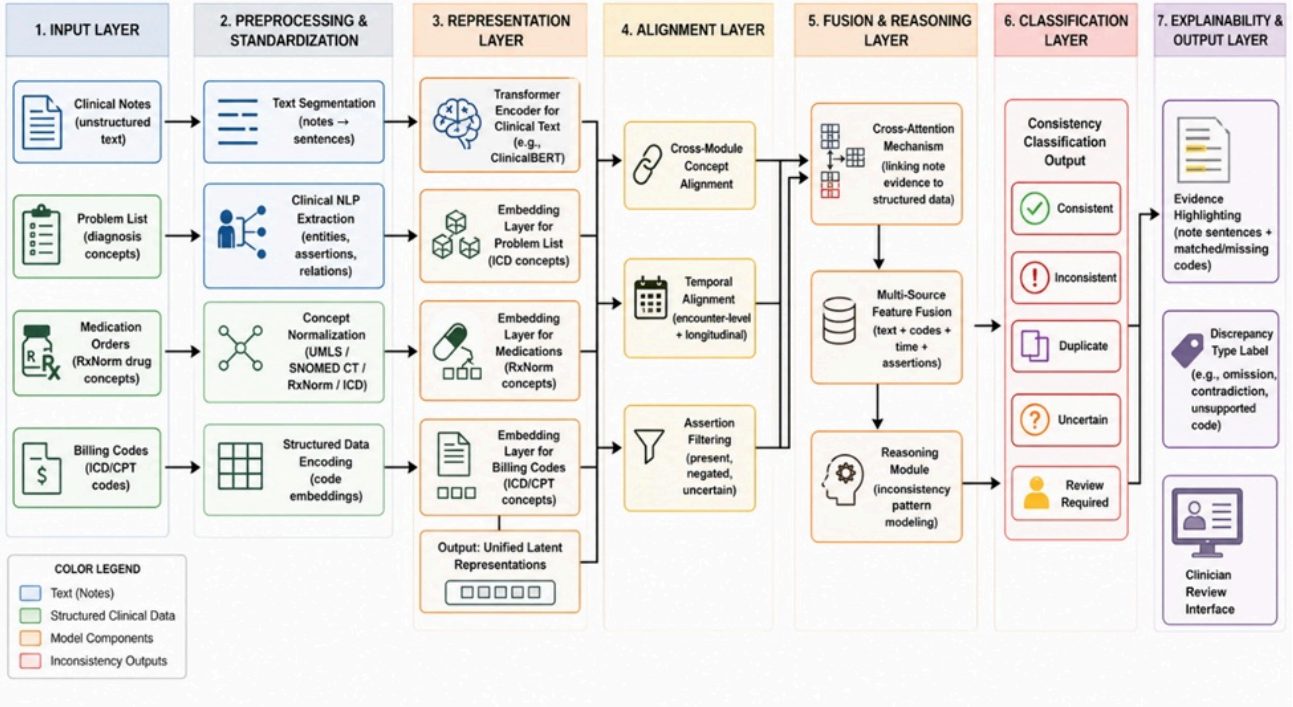


Figure 1. Hierarchical NLP architecture for cross-module inconsistency detection across clinical notes, problem lists, medication orders, and billing codes in electronic health records.

Encoder backbone for clinical text and structured data

The architecture would use a pre-trained biomedical or clinical transformer to encode note segments, with separate embedding layers for diagnosis codes, medication concepts, problem-list entries, and procedure codes. BioBERT provides a biomedical language-modeling foundation [22], while publicly available clinical BERT embeddings support representation learning over clinical narratives [23]. Structured embeddings would represent coded fields in the same latent space as narrative evidence, allowing the model to compare a note sentence with a diagnosis code, medication concept, or problem-list item. The backbone would adapt transformer-based EHR representation ideas to the specific goal of cross-module consistency analysis [24].

Cross-attention and fusion mechanism

A cross-attention module would align note sentences with problem-list entries, medication orders, and billing codes to identify which pieces of evidence support or conflict with one another. Relation extraction methods show how clinical concepts can be linked within text [12], and explainable coding models show how note evidence can be associated with administrative codes [8]. In the proposed architecture, attention could connect a sentence mentioning atrial fibrillation with an anticoagulant order and a corresponding diagnosis code, or highlight the absence of a matching problem-list item. A fusion layer would then combine textual, coded, temporal, and assertion-level representations into a joint encounter-level representation for reviewable inconsistency prediction [14, 17].

Inconsistency classification output

The final layer would classify each comparison as consistent, inconsistent, uncertain, duplicative, or requiring manual review. It could also assign a discrepancy type, such as omission, contradiction, unsupported billing code, medication–diagnosis mismatch, or outdated problem-list entry. Clinical NLP reviews caution that model outputs should be interpreted within workflow and documentation context rather than treated as automatic truth [13, 15, 25]. Therefore, the interface

should present the result as a reviewable documentation signal with supporting note sentences, structured fields, and concept mappings, not as a definitive accusation of error [20].

The functional decomposition of the proposed NLP system, including its major components and their roles in inconsistency detection, is detailed in **Table 2.**

Table 2. Functional Decomposition of the Proposed NLP Architecture for Cross-Module Consistency Analysis

Model Component	Input Type	Core Function	Technical Mechanism	Output Representation	Role in Inconsistency Detection
Clinical Text Encoder	Free-text notes	Contextual representation of narrative evidence	Transformer (ClinicalBERT/BioBERT)	Token-level embeddings	Captures clinical semantics
Structured Data Encoder	Problem list, medications, billing codes	Representation of coded clinical entities	Embedding layers	Concept vectors	Enables comparison with text
Concept Normalization Layer	Extracted entities	Map lexical variants to standard concepts	UMLS/SNOMED/RxNorm mapping	Unified concept IDs	Aligns heterogeneous data
Temporal Alignment Module	Time-stamped records	Align events across encounter timeline	Sequence modeling / timestamp matching	Time-aware embeddings	Prevents false mismatches
Cross-Attention Layer	Text + structured inputs	Identify semantic relationships across modules	Attention mechanism	Attention-weighted features	Detects support or conflict
Fusion Layer	Multi-source features	Integrate heterogeneous signals	Dense neural layers	Joint latent vector	Enables holistic reasoning
Classification Layer	Fused representation	Assign inconsistency category	Softmax classifier	Label probabilities	Produces final decision
Explanation Generator	Model outputs + inputs	Provide interpretable evidence	Attention visualization / retrieval	Highlighted text + codes	Supports clinician review

Handling Temporal and Contextual Dependencies

Temporally-aware alignment

Temporally aware alignment would link extracted concepts to the relevant phase of care rather than treating the encounter as a single static document. A medication order placed after clinical deterioration should be compared with notes and problem-list entries from the same timeframe, not only with admission documentation [4, 5]. Transformer-based

EHR representations suggest that sequential clinical information can be modeled across time [24], while clinical NLP systems can supply the note-level concepts and assertions needed for temporal comparison [1, 21]. This design would allow the model to distinguish a genuine mismatch from a documentation sequence in which evidence appears later in the encounter.

Accounting for legitimate omissions and clinical evolution

Some apparent discrepancies are clinically appropriate because diagnoses evolve, medications are discontinued, and historical conditions may remain in narrative notes without requiring active problem-list status. Clinical NLP reviews emphasize that symptoms, diseases, and contextual modifiers must be interpreted within documentation context rather than extracted as isolated strings [13, 14]. Assertion and uncertainty detection would be especially important when a note states that a condition is ruled out, suspected, resolved, or part of family history [21]. The model should therefore learn to classify some mismatches as uncertain or clinically explainable instead of treating all cross-module differences as documentation errors [15].

Handling multi-encounter longitudinal records

Longitudinal records introduce additional complexity because chronic conditions, prior procedures, and long-term medications may be documented across multiple encounters rather than repeated in every note. Large-scale EHR modeling and deep patient representations show that prior clinical history can inform interpretation of current encounters [3, 11]. A longitudinal inconsistency detector could use earlier notes, prior problem-list states, and previous medication patterns to contextualize present documentation, while still checking whether current billing and orders are supported by current evidence [4, 24]. This approach would help avoid inappropriate flags for stable chronic problems while still identifying unsupported or stale structured entries.

Model Interpretability and Clinical Audit

Generating clinician-friendly inconsistency explanations

The model should generate explanations that identify the specific evidence underlying each inconsistency flag. Explainable medical coding work demonstrates the value of linking predicted codes to note passages [8], and semantic search methods show how clinically meaningful concepts can be surfaced from narrative documentation [17]. In this proposed system, an explanation might highlight a note sentence indicating atrial fibrillation, the absence of a matching problem-list diagnosis, and the presence of an anticoagulant order. Such evidence-grounded explanations would support clinician review by making the model's reasoning inspectable rather than presenting a black-box alert [15, 25].

Audit dashboard for documentation quality

An audit dashboard could aggregate inconsistency flags by provider, department, encounter type, code family, or documentation module. Applied clinical NLP systems show how extracted information can support operational review beyond single-patient interpretation [18, 19], while documentation-oriented coding models suggest that note evidence and administrative categories can be analyzed together [7, 16]. The dashboard would not be intended to rank clinicians mechanically but to identify recurring documentation patterns requiring training, workflow redesign, or coding clarification. By presenting trends alongside examples, the system could support documentation quality improvement without reducing clinical nuance to isolated model outputs [13, 14].

Clinical Integration and Documentation Improvement

Pre-visit or post-note integration

The model could be integrated before note signing, after encounter closure, or during retrospective coding review. Real-time use would allow clinicians to reconcile problem lists, medication orders, and note content while the clinical context remains fresh, whereas retrospective use would support billing compliance and documentation audits [7, 15]. Medication extraction and note-based disease identification studies show that clinically relevant facts can be recovered from narrative text in ways that could inform downstream workflow review [5, 6, 18, 19]. To avoid excessive interruption, alerts should be limited to discrepancies with clear evidence and practical correction pathways [13, 25].

Closing the loop: from flag to correction

When a probable inconsistency is flagged, the system could suggest a documentation action such as adding a supported diagnosis to the problem list, reviewing an unsupported billing code, or clarifying a medication indication. Explainable coding and semantic search methods provide conceptual support for linking such recommendations to source evidence rather than offering unsupported suggestions [8, 17]. The correction process should remain clinician-controlled because some mismatches reflect legitimate clinical evolution, uncertainty, or local coding conventions [14, 15, 21]. A closed-loop design would therefore treat the model as a documentation assistant that accelerates review while preserving professional judgment.

Evaluation Strategy

Intrinsic NLP and consistency classification metrics

The model should be evaluated at the level of concept extraction, assertion recognition, document alignment, and inconsistency classification. Prior clinical NLP studies demonstrate the importance of assessing entity extraction, relation extraction, medication identification, and code assignment as distinct but connected subtasks [1, 2, 5-7, 12, 16]. For inconsistency classification, evaluation could consider whether the model correctly identifies omission, contradiction, duplication, unsupported code, or uncertain status without reporting speculative performance values. Because this article is conceptual, such metrics should be described as future evaluation targets rather than presented as completed results [13, 15].

Temporal validation and external generalizability

Temporal validation would examine whether a model developed on earlier encounters remains useful for later encounters after documentation workflows, coding practices, or clinical terminology change. Large EHR studies highlight the need to consider time, institution, and data-generation process when modeling clinical records [3, 4, 11, 24]. External validation would also be necessary because note templates, problem-list discipline, and billing conventions differ across hospitals and specialties. Reviews of clinical NLP emphasize that models intended for operational deployment should be tested beyond the development setting before being treated as generalizable documentation tools [13, 15, 25].

Operational impact assessment

Operational evaluation should examine whether the system improves documentation review processes without increasing clinician burden. Potential endpoints could include changes in unresolved documentation queries, coding-related denials, medication–diagnosis reconciliation workload, and time spent on manual chart review, but these should be assessed prospectively rather than assumed. Studies of note-based adverse event detection, automated coding, and condition identification show that NLP can support clinically meaningful review tasks [8-10, 18, 19]. The proposed model should therefore be evaluated not only as a classifier but as a workflow intervention affecting safety, documentation integrity, and revenue-cycle processes [7, 15, 25].

Limitations

Ambiguity in clinical language and inconsistency definition

A major limitation is that clinical inconsistency is not always objectively defined. A note may mention a suspected diagnosis, a medication may be used for multiple indications, and a billing code may reflect a clinically justified interpretation that is only indirectly documented [5, 7, 21]. Clinical NLP systems can extract concepts and assertions, but they may not fully resolve ambiguity, evolving clinical judgment, or implicit reasoning in physician documentation [13-15]. For this reason, the proposed model should support review and prioritization rather than autonomous correction.

Data quality and variation across EHR systems

EHR systems vary in note structure, problem-list maintenance, medication-order conventions, and billing-code practices. A model trained in one institution could require adaptation before use elsewhere because local documentation templates and coding workflows shape the data being modeled [3, 4]. Biomedical and clinical transformer models provide reusable language representations [22, 23], but downstream inconsistency detection would still depend on local concept mappings, workflow expectations, and review policies [17, 24]. These limitations make multi-site validation and governance essential before broad implementation [15, 25].

Conclusion

A deep learning NLP model for cross-EHR inconsistency detection could provide a unified way to compare clinical notes, problem lists, medication orders, and billing codes. By encoding narrative and structured data together, the model could identify documentation relationships that are difficult to detect through manual audit alone. Its main contribution would be to transform fragmented EHR modules into a reviewable consistency network. Such a system would support documentation integrity without replacing clinician judgment.

The key strength of the proposed model is its unified treatment of four major EHR documentation sources. Instead of examining notes, problems, medications, or billing codes in isolation, it would reason across the relationships among them. Explainable outputs would allow clinicians, coders, and quality teams to inspect the evidence behind each flag. This could improve both patient safety and revenue-cycle reliability when implemented carefully.

Important challenges remain. Clinical language is ambiguous, documentation practices vary across institutions, and not every mismatch is an error. Integration into clinical workflow would require careful attention to alert fatigue, correction authority, and medico-legal implications. The model should therefore be deployed as a review-support tool rather than an automated documentation adjudicator.

Future work should develop multi-centre validation studies and shared benchmarking corpora for EHR inconsistency detection. Such resources would allow researchers to compare model architectures, inconsistency definitions, and workflow integration strategies. A common evaluation framework would also help separate technical NLP performance from practical documentation improvement. With careful validation, deep NLP could become a meaningful component of continuous EHR data quality surveillance.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 31 Aug 2022 Revised: 26 Sep 2022 Accepted: 10 Nov 2022
Published online: 25 February 2023

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

Soysal E, Wang J, Jiang M, Wu Y, Pakhomov S, Liu H, et al. CLAMP–a toolkit for efficiently building customized clinical natural language processing pipelines. J Am Med Inform Assoc. 2018;25(3):331-6.

Si Y, Wang J, Xu H, Roberts K. Enhancing clinical concept extraction with contextual embeddings. J Am Med Inform Assoc. 2019;26(11):1297-304.

Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med. 2018;1(1):18.
https://doi.org/10.1038/s41746-018-0029-1.

Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Sci Data. 2018;5(1):180178.
https://doi.org/10.1038/sdata.2018.178.

Lin WC, Chen JS, Kaluzny J, Chen A, Chiang MF, Hribar MR, et al. Extraction of active medications and adherence using natural language processing for glaucoma patients. AMIA Annu Symp Proc. 2021;2021:773-82.

Hahn U, Oleynik M. Medical information extraction in the age of deep learning. Yearb Med Inform. 2020;29(1):208-20.
https://doi.org/10.1055/s-0040-1701975.

Teng F, Ma Z, Chen J, Xiao M, Huang L. Automatic medical code assignment via deep learning approach for intelligent healthcare. IEEE J Biomed Health Inform. 2020;24(9):2506-15.
https://doi.org/10.1109/JBHI.2020.2976906.

Mullenbach J, Wiegreffe S, Duke J, Sun J, Eisenstein J. Explainable prediction of medical codes from clinical text. In: Proc Conf North Am Chapter Assoc Comput Linguist Hum Lang Technol. 2018;1:1101-11.

Chapman AB, Peterson KS, Alba PR, DuVall SL, Patterson OV. Detecting adverse drug events with rapidly trained classification models. Drug Saf. 2019;42(1):147-56.

https://doi.org/10.1007/s40264-018-0733-7.

Jagannatha A, Liu F, Liu W, Yu H. Overview of the first natural language processing challenge for extracting medication, indication, and adverse drug events from electronic health record notes (MADE 1.0). Drug Saf. 2019;42(1):99-111.
https://doi.org/10.1007/s40264-018-0762-2.

Kataria S, Ravindran V. Electronic health records: a critical appraisal of strengths and limitations. J R Coll Physicians Edinb. 2020;50(3):262-8.
https://doi.org/10.4997/JRCPE.2020.309.

Li Z, Yang Z, Shen C, Xu J, Zhang Y, Xu H, et al. Integrating shortest dependency path and sentence sequence into a deep learning framework for relation extraction in clinical text. BMC Med Inform Decis Mak. 2019;19(Suppl 1):22.
https://doi.org/10.1186/s12911-019-0764-4.

Spasic I, Nenadic G. Clinical text data in machine learning: systematic review. JMIR Med Inform. 2020;8(3):e17984.
https://doi.org/10.2196/17984.

Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. J Am Med Inform Assoc. 2019;26(4):364-79.

Cowie MR, Blomster JI, Curtis LH, Duclaux S, Ford I, Fritz F, et al. Electronic health records to facilitate clinical research. Clin Res Cardiol. 2017;106(1):1-9.
https://doi.org/10.1007/s00392-016-1025-6.

Atutxa A, Pérez A, Casillas A. Machine learning approaches on diagnostic term encoding with the ICD for clinical documentation. IEEE J Biomed Health Inform. 2018;22(4):1323-9.
https://doi.org/10.1109/JBHI.2017.2731365.

Wu H, Toti G, Morley KI, Ibrahim ZM, Folarin A, Jackson R, et al. SemEHR: A general-purpose semantic search system to surface semantic data from clinical notes for tailored care, trial recruitment, and clinical research. J Am Med Inform Assoc. 2018;25(5):530-7.

Canales L, Menke S, Marchesseau S, D'Agostino A, del Rio-Bermudez C, Taberna M, et al. Assessing the performance of clinical natural language processing systems: development of an evaluation methodology. JMIR Med Inform. 2021;9(7):e20492.
https://doi.org/10.2196/20492.

Afzal N, Mallipeddi VP, Sohn S, Liu H, Chaudhry R, Scott CG, et al. Natural language processing of clinical notes for identification of critical limb ischemia. Int J Med Inform. 2018;111:83-9.
https://doi.org/10.1016/j.ijmedinf.2017.12.012.

Abacha AB, Mrabet Y, Sharp M, Goodwin TR, Shooshan SE, Demner-Fushman D. Bridging the gap between consumers' medication questions and trusted answers. Stud Health Technol Inform. 2019;264:25-9.
https://doi.org/10.3233/SHTI190180.

Peng Y, Wang X, Lu L, Bagheri M, Summers RM, Lu Z. NegBio: a high-performance tool for negation and uncertainty detection in radiology reports. AMIA Jt Summits Transl Sci Proc. 2018;2018:188-96.

Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics. 2020;36(4):1234-40.

Alsentzer E, Murphy JR, Boag W, Weng WH, Jin D, Naumann T, et al. Publicly available clinical BERT embeddings. In: Proc 2nd Clin Nat Lang Process Workshop. 2019:72-8.

Li Y, Rao S, Solares JRA, Hassaine A, Ramakrishnan R, Canoy D, et al. BEHRT: transformer for electronic health records. Sci Rep. 2020;10(1):7155.

https://doi.org/10.1038/s41598-020-62922-y.

Locke S, Bashall A, Al-Adely S, Moore J, Wilson A, Kitchen GB. Natural language processing in medicine: a review. Trends Anaesth Crit Care. 2021;38:4-9.
https://doi.org/10.1016/j.tacc.2021.02.007.