Journal of Health Informatics

J Health Inform. 2026;18:e1816
DOI: https://doi.org/10.59681/2175-4411.v18.iEspecial.2026.1816

ARTIGO COMPLETO
CBIS 2026

Automated extraction of billing items from clinical records using large language models

Extração automatizada de itens de faturamento em prontuários clínicos com modelos de linguagem de grande porte

Eduardo Camargo de Siqueira[1] ✉, Nilton Correia da Silva[2], Rebeca da Nóbrega Lucena Pinho[3], Maria Luisa Pontes de Assis[2], Júlia Takaki Neves[2]

ORCID: 0009-0001-2093-2783 · 0000-0002-5817-5175 · 0000-0003-3698-8287 · 0009-0005-9417-8815 · 0009-0006-5755-6925

[1] Instituto Federal de Educação, Ciência e Tecnologia de Brasília, Brasília (DF), Brasil.
[2] Universidade de Brasília, Brasília (DF), Brasil.
[3] Hospital Universitário de Brasília, Brasília (DF), Brasil.

ARTICLE INFORMATION

TYPE
Artigo completo · CBIS 2026

IDENTIFIER
e1816

DOI
10.59681/2175-4411.v18.iEspecial.2026.1816

RECEIVED
17 May 2026

ACCEPTED
20 July 2026

REVIEW PROCESS
Double Blind Review

HANDLING EDITOR
Juliano de Souza Gaspar

ACCESS
Open, CC BY 4.0

CORRESPONDING AUTHOR
Eduardo Camargo de Siqueira
eduardo.siqueira@ifb.edu.br

LICENSE

This is an open access article published under the Creative Commons CC BY 4.0 license, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited.

CC BY 4.0

ABSTRACT

Introduction: The digitalization of healthcare systems has increased the volume of unstructured clinical data, making the identification of billing items from electronic health records challenging. **Objective:** To investigate the use of large language models (LLMs) for automated billing item extraction, using the SIGTAP table as a reference. **Method:** A methodology based on text segmentation and procedure-guided inference was proposed. Four models (deepseek-r1:70b, llama3.3:70b, gemma3:27b, and qwen3:32b) were evaluated using a clinically validated dataset. **Results:** All models identify relevant procedures but consistently exhibit over-extraction. More computationally efficient models (gemma and qwen) generated larger outputs, while deepseek was more conservative and llama showed intermediate behavior. **Conclusion:** LLMs are promising as decision-support tools but still require human validation and are not suitable for fully automated hospital billing.

Keywords: Artificial Intelligence; Electronic Health Records; Hospital Billing.

RESUMO

Introdução: A digitalização dos sistemas de saúde tem ampliado o volume de dados clínicos não estruturados, dificultando a identificação de itens de faturamento em prontuários eletrônicos. **Objetivo:** Investigar o uso de modelos de linguagem de grande porte (LLMs) para a extração automatizada desses itens, utilizando a tabela SIGTAP como referência. **Método:** Foi proposta uma abordagem baseada na segmentação de textos e inferência guiada por grupos de procedimentos. A avaliação considerou quatro modelos (deepseek-r1:70b, llama3.3:70b, gemma3:27b e qwen3:32b) com base em um conjunto validado clinicamente. **Resultados:** Todos os modelos identificam procedimentos relevantes, porém apresentam tendência à superextração. Modelos mais eficientes computacionalmente (gemma e qwen) geraram maior número de itens, enquanto deepseek foi mais conservador e llama apresentou comportamento intermediário. **Conclusão:** Os LLMs são promissores como ferramentas de apoio, mas ainda requerem validação humana, não sendo adequados para automação completa do faturamento hospitalar.

Descritores: Inteligência Artificial; Registros Eletrônicos de Saúde; Faturamento Hospitalar.

CBIS 2026 XXI CONGRESSO BRASILEIRO DE INFORMÁTICA EM SAÚDE

Paper presented at the 21st Brazilian Congress of Health Informatics (CBIS 2026), held from September 23 to 25, 2026, in Brasília (DF), Brazil, and published in the CBIS 2026 section of the Journal of Health Informatics.

HOW TO CITE THIS ARTICLE: de Siqueira EC, da Silva NC, Pinho RNL, de Assis MLP, Neves JT. Automated extraction of billing items from clinical records using large language models. J Health Inform 2026;18:e1816. doi: https://doi.org/10.59681/2175-4411.v18.iEspecial.2026.1816

1 INTRODUCTION

The increasing digitalization of healthcare systems has led to a substantial growth in the volume of unstructured clinical data, particularly within electronic health records (EHRs), creating both opportunities and challenges for health service management. Among these challenges, the accurate identification of billing items from clinical documentation remains a critical issue, especially in public healthcare settings, where manual processes are time-consuming, error-prone, and often lead to underbilling and resource inefficiencies. In this context, advances in Natural Language Processing (NLP) and Large Language Models (LLMs) offer promising alternatives for transforming unstructured clinical narratives into structured, actionable information. This study proposes an automated approach for extracting billing items from clinical records using LLMs, aiming to improve the accuracy, efficiency, and scalability of hospital billing processes while contributing to more effective financial management in healthcare systems.

Recent studies highlight the rapid growth and consolidation of artificial intelligence (AI) in healthcare, expanding its applications from hospital settings to primary care, remote monitoring, and distributed care management. AI has demonstrated strong potential across clinical, administrative, and decision-support domains; however, its effective adoption depends not only on technological advances but also on organizational, methodological, and institutional factors. In clinical pharmacy, AI techniques such as machine learning, NLP, and deep learning have been applied to prescription review and error detection, although challenges remain regarding external validation and integration into clinical workflows [1]. Similarly, AI systems have achieved performance comparable to specialists in specific diagnostic tasks and contributed to reducing errors and administrative burden, including applications such as chatbots and remote monitoring, despite concerns related to privacy, bias, and user acceptance [2]. Bibliometric evidence further indicates exponential growth in AI research in healthcare, particularly in areas such as EHRs, clinical decision support, and LLMs, highlighting both rapid innovation and the need for regulatory and practical maturation [3].

AI has also shown potential to expand access to healthcare, particularly in underserved regions through telemedicine and automated triage, while wearable sensors and connected devices enable continuous real-time monitoring outside clinical environments, albeit with challenges in interoperability, standardization, and data security [4,5]. Applications in diagnosis, medical imaging, and clinical data management further reinforce the importance of data quality, workflow integration, and institutional acceptance for successful implementation [6,7]. From an organizational perspective, AI applications span diagnostic support, patient engagement, and administrative processes such as billing and revenue cycle management, though barriers to large-scale adoption are primarily related to system integration, regulation, and organizational adaptation rather than technological limitations [8]. Accordingly, governance frameworks addressing transparency, bias, data security, and lifecycle management are essential for ensuring ethical and sustainable deployment [9].

Empirical evidence demonstrates that AI can yield meaningful clinical and operational benefits, such as reduced hospital stay and mortality through early prediction models [10], improved efficiency in radiological workflows [11], and cost reductions through optimized treatment strategies [12]. However, improvements in process efficiency do not always translate into better clinical outcomes, emphasizing the importance of integration with clinical protocols and organizational structures [13]. Advances in decision-support systems have also reduced alert fatigue and improved intervention relevance [14].

Recent developments in deep learning have further advanced the modeling of longitudinal clinical data, with Transformer-based architectures such as BEHRT and Hi-BEHRT demonstrating improved performance in predicting clinical outcomes from EHR data [15]. At the same time, continual learning approaches offer potential for adaptive systems but raise challenges related to catastrophic forgetting, regulation, and ethics [16]. In the domain of clinical NLP, there has been a shift from information extraction tasks to interactive and generative applications. Transformer-based models have shown superior performance in entity recognition and classification, though limitations persist in data availability and generalization [17]. More recent work demonstrates their application in clinical chatbots and conversational systems [18], culminating in the use of LLMs and multimodal generative models capable of processing large volumes of heterogeneous clinical data. While these models enable advanced applications such as summarization and decision support, they also introduce risks related to hallucinations, bias, privacy, and regulatory compliance, reinforcing the need for robust governance and validation mechanisms [19,20].

2 METHODS

This study adopts an LLM-based approach for the automated extraction of billing items from unstructured clinical records. The proposed methodology consists of a sequential pipeline that integrates text preprocessing, document segmentation, and natural language inference, using a locally deployed model via the Ollama platform.

The input data comprise two main sources: (i) a dataset of clinical records stored in parquet format, containing record identifiers and textual clinical notes, and (ii) the SIGTAP table, provided in CSV format, which organizes billing procedures into structured groups. The clinical texts undergo a normalization step that includes removal of invalid characters, whitespace standardization, and lowercasing, ensuring consistency for downstream processing.

Given the length and complexity of clinical narratives, the texts are segmented into smaller overlapping chunks using a character-based sliding window strategy. This approach preserves contextual continuity between adjacent segments while enabling compatibility with LLM input size constraints. Each chunk is processed independently in subsequent steps.

The SIGTAP table is preprocessed to group procedures according to their organizational categories, generating subsets of candidate billing items. For each text chunk, the pipeline iteratively processes each SIGTAP group by constructing a structured prompt that includes: (i) explicit instruction constraints, (ii) the group name, (iii) a list of candidate procedures, and (iv) the clinical text chunk under analysis. The LLM is then prompted to identify, strictly based on the provided text, which items are explicitly mentioned or strongly implied, returning the output in a structured JSON format.

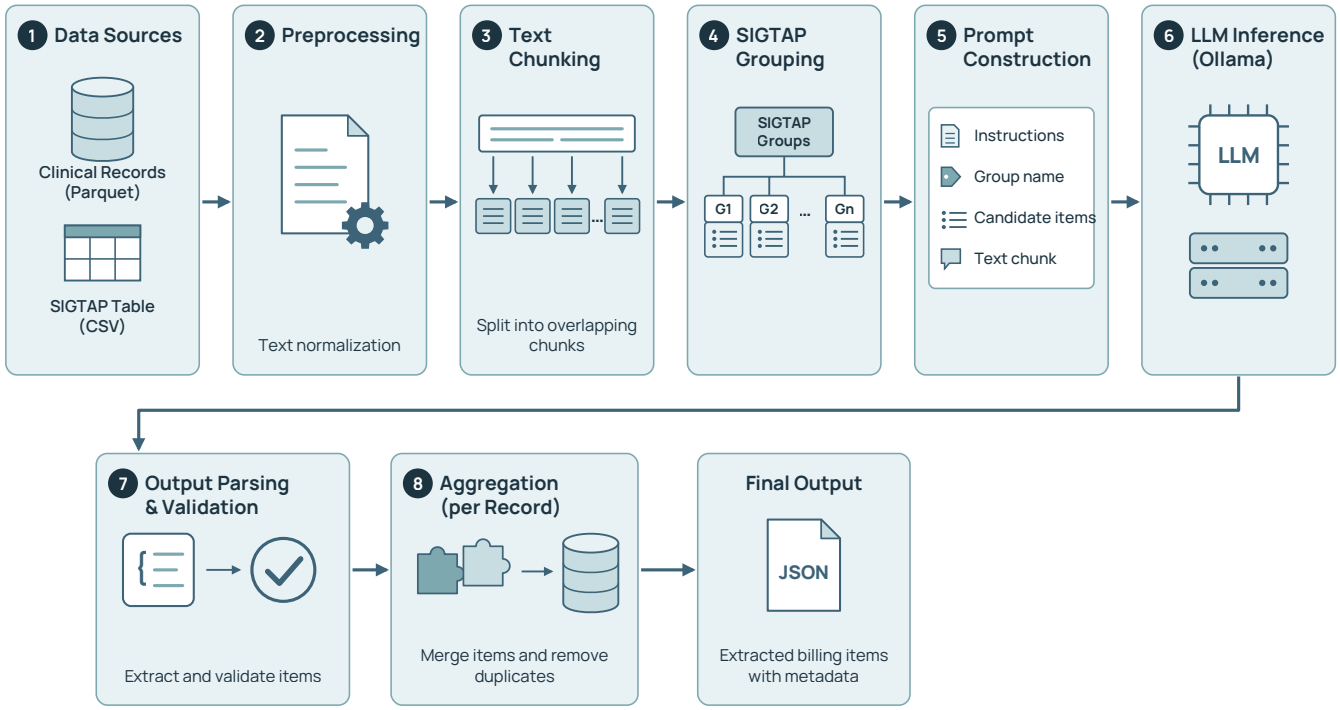
Inference is performed through calls to the Ollama API using a locally hosted deepseek model, with temperature set to zero to minimize response variability and improve consistency. To address potential inconsistencies in model outputs, a robust JSON extraction mechanism is implemented, including parsing of malformed responses and validation of returned items against the original candidate list.

Results are aggregated at the record level by combining items identified across all processed chunks and SIGTAP groups, with duplicate entries removed. Additionally, the pipeline records detailed processing metrics, including estimated input and output token counts, response time per call, total number of LLM interactions, and the number of identified items. These metrics support the evaluation of computational cost and efficiency of the proposed approach.

Finally, the extracted items are stored in a structured JSON format, including information about the clinical record, SIGTAP group, identified item, and corresponding chunk index, enabling further analysis and validation. This modular and scalable architecture allows the methodology to be applied to different clinical datasets while maintaining flexibility for adapting models and processing parameters.

Figure 1 illustrates the proposed methodology for automated extraction of billing items from clinical records using an LLM-based pipeline. The process begins with two main data sources: unstructured clinical records and the structured SIGTAP table. The clinical texts are first preprocessed and segmented into smaller chunks to enable efficient analysis. Each chunk is then combined with groups of candidate procedures from the SIGTAP table and submitted to the LLM through a structured prompt. The model identifies

Figure 1. Proposed Pipeline for Automated Extraction of Billing Items from Clinical Records



Source: Prepared by the authors (2026).

relevant billing items based solely on the textual evidence provided. Finally, the extracted items are aggregated at the record level, producing a structured output aligned with standardized billing categories. This workflow highlights the integration of data preprocessing, controlled inference, and result aggregation to support scalable and consistent information extraction.

2.1 SIGTAP Table

The SIGTAP (Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS) table is a standardized Brazilian healthcare classification system used to define and organize billing items within the Unified Health System (SUS). It comprises a structured set of procedures, medications, and medical materials, each associated with specific codes and hierarchical groupings that reflect their organizational and clinical characteristics.

In this study, the SIGTAP table is utilized in a simplified form, containing primarily two key attributes: (i) the organizational group (referred to as "*forma de organização*"), which represents a higher-level categorization of procedures, and (ii) the procedure name ("*procedimento*"), which corresponds to the specific billing item. Each group aggregates a set of semantically related procedures, enabling a hierarchical organization that facilitates systematic processing and filtering.

From a data processing perspective, the table is transformed into a dictionary-like structure in which each group is associated with a list of corresponding procedures. This representation allows the methodology to iteratively evaluate subsets of candidate billing items in a controlled manner, reducing the search space for the language model and improving extraction precision. Additionally, preprocessing steps such as normalization, deduplication, and removal of empty entries are applied to ensure data consistency and reliability.

The use of SIGTAP in this structured format is essential for constraining the extraction process, as it provides a closed set of valid billing items against which the model's outputs can be validated. This not only improves the robustness of the extraction pipeline but also aligns the results with real-world billing standards adopted in Brazilian public healthcare systems.

2.2 Clinical Records Dataset Description

The clinical records dataset used in this study consists of a subset derived from the publicly available dataset hosted in the Harvard Dataverse [21] (https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GWNBQQ). This dataset contains EHRs represented as unstructured clinical narratives, reflecting real-world documentation practices in hospital settings. Each record includes two main attributes: (i) a unique identifier ("*prontuario*"), corresponding to a patient record or hospitalization episode, and (ii) a free-text field ("*texto*"), containing detailed clinical descriptions such as patient conditions, procedures performed, medications prescribed, and diagnostic information.

As a subset of a larger dataset, the selected records preserve the inherent complexity, heterogeneity, and variability typical of clinical documentation, including differences in terminology, writing style, and level of detail. The narratives are often semi-structured and may include abbreviations, domain-specific language, and inconsistencies, which pose challenges for automated processing. Additionally, the length of the records varies considerably, ranging from short notes to extensive multi-page reports.

To support computational analysis, the dataset undergoes preprocessing steps, including text normalization and cleaning to reduce noise and ensure consistency. Due to input size limitations of large language models, each record is further segmented into overlapping text chunks, enabling efficient processing while preserving contextual continuity across segments.

The use of this dataset subset provides a realistic and representative basis for evaluating the proposed methodology, allowing the assessment of automated information extraction techniques under conditions that closely resemble practical clinical environments.

3 RESULTS

3.1 Computational Environment

All experiments were conducted in a high-performance local computing environment designed to support the execution of LLMs and the processing of large-scale clinical text data. The system is equipped with an Intel Xeon Gold 6330 processor, featuring 40 CPU cores operating at 2.00 GHz, and 150 GB of RAM, providing sufficient computational capacity for handling multiple concurrent inference tasks and large memory workloads.

The language models (deepseek-r1:70b, gemma3:27b, llama3.3:70b, and qwen3:32b) were executed locally using the Ollama framework, allowing full control over inference, data privacy, and reproducibility of experiments. This setup ensures that sensitive clinical data remains within the institutional infrastructure, addressing privacy and compliance requirements commonly associated with healthcare applications.

Overall, the computational environment provided a stable and scalable platform for evaluating the proposed methodology, demonstrating that the approach can be effectively implemented in resource-constrained yet realistic settings without reliance on external cloud-based services.

Table 1. Agreement between models by clinical record

record	deepseek	llama	gemma	qwen	agreement	union
1	5	9	5	9	3	15
2	23	34	40	31	12	68
3	17	22	26	24	9	49
4	13	20	22	21	7	42
5	7	18	30	19	4	44
6	15	19	20	18	8	39
7	18	27	38	29	10	61
Total	98	149	181	151	53	318

Source: Prepared by the authors (2026).

Table 2. Comparison between extracted items and clinically valid items

record	ground truth	deepseek / gt	llama/ gt	gemma / gt	qwen / gt
1	5	1.0×	1.8×	1.0×	1.8×
2	5	4.6×	6.8×	8.0×	6.2×
3	5	3.4×	4.4×	5.2×	4.8×
4	4	3.2×	5.0×	5.5×	5.3×
5	4	1.8×	4.5×	7.5×	4.8×
6	3	5.0×	6.3×	6.7×	6.0×
7	3	6.0×	9.0×	12.7×	9.7×

Source: Prepared by the authors (2026).

3.2 Clinical Review of Extracted Billing Items

A qualitative clinical review was conducted to assess the billing items extracted by four large language models from annotated clinical discharge summaries. The evaluation considered whether each suggested SIGTAP item was explicitly documented or clinically supported by the content of the records.

Overall, the models exhibited distinct extraction behaviors. Specifically, gemma3:27b remained the most expansive model, generating the highest number of candidate items, while deepseek-r1:70b maintained a more conservative profile. The newly introduced models showed intermediate behaviors: llama3.3:70b tended toward a balanced extraction pattern, whereas qwen3:32b demonstrated a slightly more permissive inference strategy, particularly in complex clinical contexts.

Across the seven analyzed records, substantial variability was observed not only in the number of extracted items but also in the overlap between models. Agreement levels remained low across all combinations, reinforcing that model architecture and training data substantially influence extraction outcomes.

Table 1 indicates that llama3.3:70b consistently retrieves a broader set of candidate billing items, albeit at the cost of increased need for clinical validation. In contrast, deepseek-r1:70b remains more conservative, prioritizing precision over recall. The overall low agreement across models highlights that model selection has a direct and substantial impact on both the quantity and nature of suggested billing items.

Table 2 presents the comparison between extracted billing items and the clinically validated ground truth across anonymized records. The results confirm a consistent pattern of over-extraction across all evaluated models, though with varying intensity.

deepseek-r1:70b demonstrates moderate inflation, ranging from 1.0× to 6.0× the number of clinically valid items, suggesting that it captures relevant procedures while limiting excessive inference. In contrast, llama3.3:70b exhibits the highest level of overestimation, reaching up to 12.7× the ground truth, indicating a strong tendency to infer procedures not explicitly documented.

The gemma3:27b model presents a more controlled inflation pattern, generally closer to deepseek, while qwen3:32b demonstrates a higher expansion rate, often approaching llama-like behavior, especially in complex cases.

Notably, all models perform accurately in simpler records (e.g., Record 1), where procedures are explicitly described. However, as clinical complexity increases, divergence from the ground truth becomes more pronounced, reinforcing the limitations of relying solely on LLMs for billing extraction.

In summary, the results indicate that the proposed LLM-based approach is effective for generating candidate billing items from clinical narratives, particularly in cases where procedures are clearly documented. However, all evaluated models exhibit varying degrees of over-extraction, with a consistent trade-off between sensitivity and precision.

Table 3. Comparative computational performance of LLMs across clinical records

model	avg. time per call (s)	total tokens (avg.)	calls per record (avg.)	total time (avg., s)	items extracted (avg.)
deepseek-r1:70b	3.25	233,194	296	946.27	14.00
llama3.3:70b	2.31	233,771	296	690.48	21.29
gemma3:27b	1.40	234,216	296	416.81	25.86
qwen3:32b	1.26	233,412	296	375.98	21.57

Source: Prepared by the authors (2026).

Therefore, the outputs should be interpreted as a decision-support layer rather than an autonomous billing mechanism. A human validation step remains essential, especially to distinguish explicitly documented procedures from clinically plausible but unsupported inferences.

3.3 Computational Results

This subsection presents a comparative analysis of the computational performance of four LLMs applied to the proposed billing item extraction pipeline.

From a computational perspective, qwen3:32b and gemma3:27b demonstrated the highest efficiency, with average response times of 1.26 and 1.40 seconds per call, respectively. These models achieved substantially lower total execution times, averaging 375.98 and 416.81 seconds per record. In contrast, llama3.3:70b and deepseek-r1:70b presented higher latency, with average times of 2.31 and 3.25 seconds per call, respectively, resulting in total execution times of 690.48 and 946.27 seconds.

Table 3 summarizes the comparative computational performance of the evaluated models. While token usage and number of calls remained consistent across models, notable differences emerged in execution time and extraction output.

Regarding extraction behavior, a clear trade-off emerges between computational efficiency and output expansion. The gemma3:27b model produced a substantially larger number of extracted items, averaging 25.86 items per record. This indicates an extremely expansive inference behavior, suggesting high recall but likely low precision. qwen3:32b also demonstrated high extraction capacity, with an average of 21.57 items per record.

In contrast, llama3.3:70b and deepseek-r1:70b produced fewer items, with averages of 21.29 and 14.00, respectively. Among them, deepseek-r1:70b remained the most conservative model, also yielding the lowest number of unique items.

These results reveal a strong inverse relationship between computational cost and extraction behavior. More efficient models (gemma and qwen) tend to generate substantially larger candidate sets, while more computationally intensive models (deepseek) produce more restrained outputs.

4 DISCUSSION

The results of this study provide important insights into the capabilities and limitations of LLMs for extracting billing items from unstructured clinical records. While all evaluated models demonstrated the ability to identify clinically relevant procedures, the analysis revealed a consistent pattern of over-extraction, particularly in more complex records. This behavior becomes even more pronounced when comparing models with different architectures and inference strategies.

A key finding is that model behavior is influenced not only by explicit textual evidence but also by implicit clinical reasoning learned during training. This is especially evident in the performance of llama3.3:70b, gemma3:27b, and qwen3:32b, which frequently inferred procedures based on diagnoses, comorbidities, or typical care pathways, even when such procedures were not explicitly documented. Among these, gemma3:27b exhibited the most expansive behavior, generating substantially larger sets of candidate items, while qwen3:32b and llama3.3:70b also demonstrated high levels of contextual inference. From a clinical and administrative perspective, this introduces a substantial risk of overbilling, which may lead to compliance issues in regulated healthcare systems such as the Brazilian SUS.

In contrast, deepseek-r1:70b exhibited a more conservative extraction profile, producing fewer items and showing reduced levels of overestimation. However, this behavior comes at the cost of under-extraction, where clinically valid billing items may be missed. The comparison across models reveals a clear trade-off between sensitivity (recall) and specificity (precision), which is central to the deployment of LLMs in this domain. While gemma3 and qwen3 favor recall at the expense of precision, and llama3 occupies an intermediate position, deepseek provides a more restrained (though still imperfect) approximation.

From a computational perspective, the results also highlight an important trade-off between efficiency and extraction behavior. More computationally efficient models, such as gemma3:27b and qwen3:32b, tend to produce substantially larger sets of candidate items, suggesting that reduced computational constraints may allow broader exploration of possible

outputs. In contrast, more computationally intensive models, such as deepseek-r1:70b, generate more conservative outputs. This indicates that model selection impacts not only inference cost but also the semantic scope of extraction.

From an operational standpoint, these results indicate that current LLM-based approaches are not yet suitable for fully automated billing systems. Instead, their most appropriate role lies in decision-support scenarios, where they assist human auditors by generating candidate billing items. In this context, a hybrid workflow can be envisioned: a high-recall model (such as gemma3 or qwen3) generates a broad set of candidates, followed by filtering through a more conservative model (such as deepseek) or rule-based validation, and final confirmation by a healthcare professional.

Overall, the findings reinforce that LLMs should be integrated as assistive tools within structured validation pipelines, rather than deployed as autonomous systems. Future work should focus on improving alignment between model outputs and clinical documentation standards, reducing spurious inference, and incorporating domain-specific constraints to enhance reliability in real-world healthcare applications.

5 CONCLUSIONS

This study investigated the use of LLMs for the automated extraction of billing items from unstructured clinical records, using SIGTAP as the reference framework. The results demonstrate that LLMs are capable of identifying clinically relevant procedures and transforming free-text narratives into structured candidates for billing. This confirms the potential of generative AI as a supporting tool for improving efficiency in hospital billing processes.

However, the evaluation against a clinically validated ground truth revealed a consistent pattern of over-extraction, particularly in more complex records. Models tend to infer procedures based on clinical context rather than explicit documentation, leading to a substantial number of false positives. At the same time, more conservative models exhibit under-extraction, missing relevant billable items. This highlights an inherent trade-off between recall and precision in LLM-based approaches.

From a practical perspective, these findings indicate that LLMs are not yet suitable for fully autonomous billing systems. Instead, their most appropriate role lies in decision support, where they assist healthcare professionals by suggesting candidate items that must be validated before submission. In this context, hybrid approaches combining multiple models and incorporating human-in-the-loop validation are particularly promising.

In addition, the study underscores the importance of structured constraints and post-processing mechanisms, such as alignment with standardized coding systems (e.g., SIGTAP), to improve reliability and reduce inappropriate inferences. The dependency on explicit procedural documentation further suggests that improvements in clinical recording practices could directly enhance the performance of AI-based extraction systems.

Author contributions. The authors contributed equally to all stages of the article: conception and design of the study; data collection, analysis, and interpretation; drafting and critical revision of the manuscript; and approval of the final version.

Acknowledgments. The authors thank the University of Brasília (UnB), the Federal Institute of Education, Science and Technology of Brasília (IFB), University Hospital of Brasília (HUB), and the Samsung Ocean Brasil for the institutional and academic support that enabled the development of this research. They also thank the Research Support Foundation of the Federal District (FAP-DF) for the financial support, which was essential for conducting the research activities.

Funding. This work was supported by the Research Support Foundation of the Federal District (FAP-DF).

Data and materials availability. The authors did not provide a data and materials availability statement.

Conflicts of interest. The authors did not declare any conflicts of interest.

Use of artificial intelligence. In this work, Generative Artificial Intelligence tools were used to support grammatical, orthographic, and stylistic revision. Their use aimed to improve the clarity, cohesion, and overall textual quality of the manuscript, without any interference in the scientific content, data interpretation, methodology, or presented results. It is emphasized that all analyses, methodological decisions, and conclusions are the sole responsibility of the authors.

REFERENCES

1. Ranchon F, Chanoine S, Lambert-Lacroix S, Bosson JL, Moreau-Gaudry A, Bedouch P. Development of artificial intelligence powered apps and tools for clinical pharmacy services: A systematic review. Int J Med Inform. 2023;172:104983. doi: https://doi.org/10.1016/j.ijmedinf.2022.104983

2. Singareddy S, Sn VP, Jaramillo AP, Yasir M, Iyer N, Hussein S, et al. Artificial Intelligence and Its Role in the Management of Chronic Medical Conditions: A Systematic Review. Cureus. 2023;15(9):e46066. doi: https://doi.org/10.7759/cureus.46066

3. Xie Y, Zhai Y, Lu G. Evolution of artificial intelligence in healthcare: a 30-year bibliometric study. Front Med (Lausanne). 2024;11:1505692. doi: https://doi.org/10.3389/fmed.2024.1505692

4. Lamem MFH, Sahid MI, Ahmed A. Artificial intelligence for access to primary healthcare in rural settings. Journal of Medicine, Surgery, and Public Health. 2025;5:100173. doi: https://doi.org/10.1016/j.glmedi.2024.100173

5. Chen S, Fan S, Qiao Z, Wu Z, Lin B, Li Z, et al. Transforming Healthcare: Intelligent Wearable Sensors Empowered by Smart Materials and Artificial Intelligence. Adv Mater. 2025;37(21):e2500412. doi: https://doi.org/10.1002/adma.202500412

6. Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya AI, Almohareb SN, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ. 2023;23(1):689. doi: https://doi.org/10.1186/s12909-023-04698-z

7. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56. doi: https://doi.org/10.1038/s41591-018-0300-7

8. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J. 2019;6(2):94-8. doi: https://doi.org/10.7861/futurehosp.6-2-94

9. Hassan M, Borycki EM, Kushniruk AW. Artificial intelligence governance framework for healthcare. Healthc Manage Forum. 2025;38(2):125-30. doi: https://doi.org/10.1177/08404704241291226

10. Shimabukuro DW, Barton CW, Feldman MD, Mataraso SJ, Das R. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. BMJ Open Respir Res. 2017;4(1):e000234. doi: https://doi.org/10.1136/bmjresp-2017-000234

11. Arbabshirani MR, Fornwalt BK, Mongelluzzo GJ, Suever JD, Geise BD, Patel AA, et al. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. NPJ Digit Med. 2018;1:9. doi: https://doi.org/10.1038/s41746-017-0015-z

12. Voermans AM, Mewes JC, Broyles MR, Steuten LMG. Cost-Effectiveness Analysis of a Procalcitonin-Guided Decision Algorithm for Antibiotic Stewardship Using Real-World U.S. Hospital Data. OMICS. 2019;23(10):508-15. doi: https://doi.org/10.1089/omi.2019.0113

13. Connell A, Montgomery H, Martin P, Nightingale C, Sadeghi-Alavijeh O, King D, et al. Evaluation of a digitally-enabled care pathway for acute kidney injury management in hospital emergency admissions. NPJ Digit Med. 2019;2:67. doi: https://doi.org/10.1038/s41746-019-0100-6

14. Segal G, Segev A, Brom A, Lifshitz Y, Wasserstrum Y, Zimlichman E. Reducing drug prescription errors and adverse drug events by application of a probabilistic, machine-learning based clinical decision support system in an inpatient setting. J Am Med Inform Assoc. 2019;26(12):1560-5. doi: https://doi.org/10.1093/jamia/ocz135

15. Li Y, Mamouei M, Salimi-Khorshidi G, Rao S, Hassaine A, Canoy D, et al. Hi-BEHRT: Hierarchical Transformer-Based Model for Accurate Prediction of Clinical Events Using Multimodal Longitudinal Electronic Health Records. IEEE J Biomed Health Inform. 2023;27(2):1106-17. doi: https://doi.org/10.1109/JBHI.2022.3224727

16. Lee CS, Lee AY. Clinical applications of continual learning machine learning. Lancet Digit Health. 2020;2(6):e279-e281. doi: https://doi.org/10.1016/S2589-7500(20)30102-3

17. Yin J, Ngiam KY, Teo HH. Role of Artificial Intelligence Applications in Real-Life Clinical Practice: Systematic Review. J Med Internet Res. 2021;23(4):e25759. doi: https://doi.org/10.2196/25759

18. Babu A, Boddu SB. BERT-Based Medical Chatbot: Enhancing Healthcare Communication through Natural Language Understanding. Explor Res Clin Soc Pharm. 2024;13:100419. doi: https://doi.org/10.1016/j.rcsop.2024.100419

19. Fahim YA, Hasani IW, Kabba S, Ragab WM. Artificial intelligence in healthcare and medicine: clinical applications, therapeutic advances, and future perspectives. Eur J Med Res. 2025;30(1):848. doi: https://doi.org/10.1186/s40001-025-03196-w

20. Chong PL, Vaigeshwari V, Mohammed Reyasudin BK, Noor Hidayah RA, Tatchanaamoorti P, Yeow JA, et al. Integrating artificial intelligence in healthcare: applications, challenges, and future directions. Future Sci OA. 2025;11(1):2527505. doi: https://doi.org/10.1080/20565623.2025.2527505

21. Silva RPD, Pazin-Filho A. Dataset of anonymized discharge summaries of sepsis patients from a Brazilian tertiary hospital for NLP applications. Data Brief. 2025;61:111804. doi: https://doi.org/10.1016/j.dib.2025.111804

EDITORIAL NOTES

The Journal of Health Informatics is a scientific journal maintained by the Brazilian Society of Health Informatics (Sociedade Brasileira de Informática em Saúde), a non-profit scientific association dedicated to the development of Health Informatics and Digital Health in Brazil. The editorial process is conducted with scientific independence, free from institutional, commercial, or sponsor interference, and follows international best practices in publication ethics.

This article was double-blind peer reviewed by specialists in the field. The editorial team used artificial intelligence tools, under human supervision, in the initial screening, in checking the text, references and numbers and, when needed, in redrawing figures, charts and tables, which the authors approved in the proof. The tools and safeguards are described in the journal's Policy on the use of artificial intelligence. Editorial decisions were made by people.

SBIS Sociedade Brasileira de Informática em Saúde