ISSN: 2754-6659

Journal of Artificial Intelligence & Cloud Computing

SCIENTIFIC Research and Community

Review Article

Open Access

Addressing Clinical Documentation Challenges: The Role of AI-Powered Digital Scribes

Wasim Fathima Shah

Independent Researcher, USA

ABSTRACT

In the realm of healthcare, the burden of clinical documentation has long been a challenge for healthcare professionals. This paper presents a comprehensive study on the development and evaluation of a unified digital scribe system aimed at automating the process of recording, transcribing, and annotating clinical narratives in Spanish. The proposed system integrates cutting-edge technologies, including speech recognition and Named Entity Recognition (NER), to streamline the documentation workflow.

The study evaluates the performance of the system in a simulated environment, encompassing transcription accuracy and entity recognition across medical and dental domains. Notably, the average Word Error Rates (WER) for different domains reveal promising results, showcasing the system's potential to accurately transcribe spoken clinical narratives. Furthermore, the NER model demonstrates its capability to identify key medical entities such as diseases, body parts, and medications, with F1 scores indicating strong performance.

The paper also outlines the development of a prototype platform that unifies both the transcription and NER modules. This platform offers healthcare professionals a seamless solution for capturing, transcribing, and annotating clinical information, ultimately aiming to reduce documentation burden and improve clinical workflow efficiency.

In conclusion, this research presents a robust foundation for the development of a digital scribe system in the Spanish healthcare system, offering promising solutions to alleviate clinical documentation challenges, reduce professional burnout, and enhance patient care through streamlined documentation practices.

***Corresponding author**

Wasim Fathima Shah, Independent Researcher, USA.

Received: January 07, 2022; **Accepted:** January 14, 2022; **Published:** January 22, 2022

Keywords: Clinical Documentation, Digital Scribe, Speech Recognition, Named Entity Recognition (NER), Healthcare Automation, Spanish Language, Medical Transcription, Clinical Narratives, Documentation Burden, Healthcare Efficiency

Introduction

Electronic Health Records (EHR) hold crucial data pertaining to medical consultations, encompassing examinations, prescriptions, and hospitalizations. These records serve not only administrative and legal functions but also play a pivotal role in aiding clinical decision-making and enhancing the secondary utilization of medical information. The acquisition of these records predominantly involves healthcare professionals who split their time between patient interaction and computer usage [1].

In the context of a healthcare appointment, various stages occur. The initial minutes entail greetings between the doctor and the patient, along with the registration of the patient's personal details. Subsequently, anamnesis follows, involving the collection of information regarding the present ailment, including the reason for the consultation and prevailing symptoms, as well as obtaining remote data about previously diagnosed diseases and family medical history to discern potential correlations with the current condition and its treatment.

Next, the physical examination of the patient is conducted, a pivotal step in identifying signs of a potential illness that might be affecting the patient, thus elucidating the reason for the consultation. Subsequent to this, supplementary examinations may be deemed necessary. Finally, the diagnosis, its corresponding treatment, and prognosis are provided. This concluding stage exhibits variability, contingent upon the identified pathologies, with outcomes ranging from simple prescriptions to surgical interventions [2].

Upon concluding the consultation, the record is documented through the process of evolution and discharge summary, capturing essential findings [3]. All these records hold the potential to be beneficial for future medical care, serving not only administrative and legal purposes but also acting as reference points for novel techniques, procedures, and epidemiological studies.

The tools employed for composing and storing a patient's clinical record have undergone significant advancements in recent decades, often replacing traditional paper-based methods with electronic ones [4]. However, the digital records are frequently underutilized due to a lack of awareness regarding their potential or the necessity for additional tools and knowledge to preprocess them.

It is estimated that over 40% of the data within electronic clinical records consist of unstructured text. This poses challenges for analysis due to the widespread use of non-standardized abbreviations, variability in clinical language across medical specialties and healthcare professionals, and limitations on data accessibility for privacy reasons, among other factors [1].

Furthermore, several challenges and issues are associated with the quality of data obtained through clinical documentation systems:

Interference in Doctor-Patient Relations

Healthcare professionals' involvement in clinical and administrative processes during consultations can compress the time dedicated to patient care, hindering effective clinical interaction. Some professionals simultaneously document while conducting clinical actions, impeding communication. Additionally, institutions allocating time for documentation later in the workday may result in lower-quality documentation [5].

Transcription Quality Issues

Time constraints in the clinical documentation process may lead to increased typing speed, resulting in misspelled words and the use of confusing abbreviations, potentially leading to medical errors in patient care [6].

Usability Limitations of Documentation Interfaces Due To Cross-Contamination

The use of human interaction devices like keyboards and mice during the care process can increase the risk of Healthcare-Associated Infections (HAIs), especially in dental care, which generates aerosols that can contaminate the environment [7].

These problems collectively contribute to healthcare professionals' exhaustion, negatively impacting cognitive capacity and increasing the risk of errors and burnout [8]. To address these issues, digital scribe systems have emerged as a promising solution. These systems can capture information dictated by healthcare professionals, automatically generating documentation similar to human medical scribes. Implementing such systems allows professionals to focus more on patient care, enhancing communication, reducing documentation time, increasing productivity, and mitigating burnout [5]. However, it's worth noting that the available digital scribe solutions primarily support clinical documentation processes in English, with no known options for Spanish that offer transcription of clinical audio and key information detection.

To tackle these challenges, this work presents a comprehensive and automated digital scribe designed to transcribe clinical audio notes in Spanish. Additionally, it incorporates a clinical entity recognition model to extract pertinent information that can aid in treatment decision-making. This approach enables the generation of a summary of a patient's history without the need to read the entire document. Furthermore, the study conducts an in-depth linguistic analysis to evaluate errors introduced by speech-to-text recognition systems and outlines the deployment of this tool as an application.

Background

Our proposed system consists of two primary modules: the speech-to-text recognition module and the clinical information extraction model. These components will be examined separately.

Automatic speech recognition (ASR)The speech-to-text system, also known as ASR, aims to extract acoustic information from audio recordings and convert this audio content into a sequence of words or lexical units [9,10]. At a higher level, the ASR system comprises two primary components: an acoustic model and a language model. The acoustic model encompasses the acoustic front-end, which extracts spectral features of the audio on a segment-by-segment basis, with the intention of distinguishing the fundamental language units, known as phonemes [11].

The language model, on the other hand, incorporates higher-level knowledge about the mapping from words to phonemes (the dictionary) and the grammatical rules of the language. Training the ASR system's parameters requires annotated data from the application domain, which forms the basis of the training process. To achieve accurate speech-to-text performance, it is essential to record audio files with high temporal resolution and amplitude quality, alongside a sufficient volume of annotated data for developing representative models for the task. ASR systems are traditionally evaluated using the word error rate (WER), a metric that quantifies substitution, insertion, and omission errors [10].

Currently, various commercial alternatives for ASR systems have been developed, including Microsoft Speech API (SAPI), Amazon Transcribe, IBM Watson Speech-to-Text, Google Cloud Speech-to-Text API, and Wit [12-14].

Information Extraction

The second challenge addressed by our digital scribe system involves the extraction of clinically relevant information from text transcriptions, which occurs after the speech-to-text transcription process. Common techniques for retrieving crucial text information primarily rely on the field of Natural Language Processing (NLP). In particular, the NLP task used to address this issue is Named Entity Recognition (NER), which focuses on extracting text segments referencing predefined categories. Notably, this task presents challenges, particularly when applied to automatically transcribed text, as traditional NER models tend to perform better on grammatically coherent text with clearly defined sentences. These features can be compromised in automatically transcribed text due to the lack of punctuation marks like periods and commas, leading to incoherent structure.

NER requires a labeled dataset with annotated entities for training. In recent years, several datasets specific to the Spanish clinical domain have been released, including the IXAMed corpus for identifying diseases and drugs, the DrugSemantics corpus for recognizing multiple entity types, including diseases, and Clinical Trials annotated with chemical, anatomical, procedural, and disorder mentions [15-17]. Other datasets, such as DISTEMIST, focused on disease mention recognition and normalization, and the Chilean Waiting List corpus, consisting of referrals from the Chilean public healthcare system, have been created for shared task purposes [18-20].

The freely available models of the Chilean Waiting List Corpus played a pivotal role in our research [21]. Our primary objective was to identify diseases, medications, and body parts related to patients. The expected performance of these models ranged between an F1 score of 0.83 for disease recognition, 0.84 for medication recognition, and 0.87 for body parts recognition.

Data and Methods

Proposed System

ASR Module

For the speech-to-text transcription step, we employed the Google Cloud Speech-to-Text service, known for its high performance and cost-effectiveness. Specifically, we used the version optimized for short audio, enabling the processing of up to one minute of voice audio data. In the ASR process, data is sent either synchronously or asynchronously, processed, and the resulting transcription is returned. Response times vary based on audio file length and quality. We incorporated the API from Google Cloud and made modifications to generate text files for subsequent stages.

Named Entity Recognition (NER) Module

Regarding the clinical NER model, we followed the same architecture proposed by Rojas et al. [21]. This model comprises three primary submodules: the embedding layer, the encoding layer, and the decoding layer. Initially, it utilizes static and contextual word embeddings derived from Chilean waiting list referrals to encode sentences. The output is then passed to a Bidirectional Long Short-Term Memory (BiLSTM) encoding layer to capture long-contextual information about the tokens within the current sentence. Finally, the model employs a Conditional Random Field (CRF) layer and the Viterbi decoding algorithm to recognize clinical entities within the current sentence. The code for this model is made available to the research community [1].

Speech Recognition Experiments

To evaluate the performance of the speech recognition module, we had four participants read three sets of texts, which constituted the gold standard. From this gold standard, we selected 60 texts from Chilean public healthcare system referrals and 30 texts of similar length from the general domain. Among the 60 clinical texts, 30 were from dental consultations and 30 from medical referrals. We standardized these referrals to facilitate analysis and comparison, expanding abbreviated words for proper reading. In summary, the gold standard for evaluating the transcription of 90 standardized texts included 30 dental, 30 medical, and 30 general-domain texts. The participants recorded audio using Audacity® software, saving the files in uncompressed .wav format with a frequency of 44100 Hz and 16-bit linear quantization (PCM) [2]. They used hands-free headphones with built-in microphones and wore triple-layer masks due to the COVID-19 pandemic. Recordings were made in a relatively quiet environment, and any recordings with a duration exceeding one minute or a size larger than 10 MB were excluded, as they exceeded the service's transcription limit. Four participants, including a dental surgeon, accountant, law student, and dental surgeon, each read all 90 texts.

Transcription performance was evaluated using the Word Error Rate (WER), a metric that quantifies discrepancies between the ASR system's word sequence and the gold standard. We used the open-source Python module "asr-evaluation" for this purpose, which automatically classifies machine translation output errors [3].

Manual Annotation for NER

Manual annotation involved humans identifying predefined pieces of information according to annotation guidelines and assigning corresponding labels. To ensure quality, multiple annotators were employed, and intra/inter annotator agreement was monitored. In this section, only medical and dental texts were annotated with clinical entities, as general-domain texts were used solely for comparing transcription quality between general and clinical domains.

A team of three researchers annotated the 60 medical-dental domain texts, collaborating with two healthcare professionals to reach a consensus on the annotations made in the texts [19]. The NER model was trained using annotations from the Chilean Waiting List corpus, allowing it to recognize multiple entities [21]. Three main entity types were focused on: Diseases, Body Parts, and Medications.The annotation results were in standoff format, including annotations of tokens with multiple labels, known as nested entities [4, 20].

The manual annotations were compared with the automatic annotations found in the original and transcribed texts, evaluating whether the Google Cloud Speech-to-Text service had any impact on annotations when used in conjunction with other tools. The comparison was measured using the F1 score, a standard metric for assessing agreement between NER systems and gold standard sets. All code, audio recordings, transcriptions, and annotations are freely available to facilitate replication of our experiments [5]. In the following section, we delve into the error analysis of each module.

Results

To gain a deeper understanding of the limitations of the digital scribe, we conducted an error analysis of both components of our system.

Transcription Errors

A total of 360 recordings were distributed among the four narrators (90 each) and subsequently transcribed. Plain text files were obtained and compared with their respective gold standard files. The distribution of the 360 WER values was plotted, and a Shapiro-Wilk test was performed to assess their normality. The test returned a p-value of 0.017, indicating non-normal distribution of the dataset.

The average WER values for the three domains are as follows: 10.44% for dental domain transcripts, 9.98% for medical transcripts, and 9.06% for general domain transcripts. Comparing the dental-medical domain and general-domain audio, we observe a higher WER in the former.

We also conducted a Kruskal-Wallis test to compare the means of the three groups (dental, medical, and general) since they are independent groups. The test resulted in a p-value of 0.010, indicating a significant difference in the means of at least one of the three groups. Additionally, Mann-Whitney tests were performed to compare the means of the two groups, and the obtained p-values are as follows: 0.006 (medical vs. general), 0.0512 (dental vs. general), and 0.199 (dental vs. medical). The symbol ** indicates a significance value less than 0.05, while the symbol * indicates a greater value.

Furthermore, a Spearman correlation was used to explore a potential relationship between text length and WER values, yielding a p-value of 0.08, supporting the null hypothesis that there is no such relationship. We also compared means among different narrators using the Wilcoxon test for related samples, resulting in a p-value of 0.088, indicating no significant differences among the narrators.

Analysis of Linguistic Errors in Transcriptions

During the speech-to-text transcription process, a total of 1,152 errors were identified and classified into three main categories:

Lexical Level or Word Form Errors

Substitution: Arbitrary replacement of one lexical unit with another, often with similar phonetics. Insertion and deletion: Arbitrary inclusion or omission of lexical units, especially prepositions, affecting syntactic structure.

Morphological Level or Internal Word Structure Errors
Number concordance errors due to dialectal characteristics: Plural identification challenges, particularly in Chilean Spanish, where /s/ is aspirated at the end of syllables, impacting plural recognition and concordance in syntactic structures.

Gender concordance errors due to phoneme-to-grapheme discrepancies: Gender is a formal marker that, in certain instances, correlates with characteristics of sexed animate entities. In Spanish, gender is manifested through morphemes, with the suffixes ✯✯-e✯✯ and ✯✯-o✯✯ representing the masculine gender, and ✯✯-a✯✯ representing the feminine gender. Additionally, the masculine form serves as the unmarked gender. In the corpus, we observe arbitrary changes in the gender of nouns, impacting the concordance within syntactic structures. These changes could lead to subsequent alterations in grammatical categories and semantic shifts, as illustrated in examples (8), or modifications in verb forms, as seen in (9):

restos dentales (masculine) => rest✿as dentales (feminine)

paciente refiere que la madre present ✿a episodios (masculine) => paciente refiere que la madre presente episodios (feminine)

Spelling or Graphic Conventions Errors
This category of errors encompasses three cases, including accentuation, punctuation, and arbitrary spacing, all of which constitute issues related to typographical syntax. Accentuation errors, in particular, can alter the grammatical category of a word, as demonstrated in examples (10) and (11):

soplo card´ıaco no especific✿o y diabetessoplo card´ıaco no especific´o y diabetes
alt. digestivas, distensi´on abdominal, solicit✿o evaluaci´onalt. digestivas, distensi´on abdominal, solicit´o evaluaci´on

Critical Errors for Clinical Practice
Lexical substitution errors fall into this category and can potentially mislead clinicians reviewing the records. Examples (1), (2), and (3) illustrate cases where replacing one word with another that is phonetically similar leads to a change in meaning within a specialized vocabulary context. However, clinicians can often disambiguate such cases based on the sentence context, and these errors may not necessarily have a critical impact on clinical practice.

Another aspect of transcription revision involves instances where the system successfully corrects lexical or syntactic inadequacies present in the original text, as shown in (12):

con antecedentes de abscesos submucosos en pieza_ 7con antecedentes de abscesos submucosos en piezas 7

In this example, the system rectifies a grammatical error related to the word ✯✯pieza✯✯, which should be plural in this context, and makes the necessary correction during transcription.

NER Performance
The second analysis assessed the NER model's performance on both the original text and transcriptions generated from each narrator's dictation. The gold standard for comparison was the annotations manually created by three healthcare professionals, limited to medical and dental domain texts. Table 2 displays the primary results, with columns indicating the NER model's performance on original texts, as well as transcriptions from four different narrators.

To compare the groups and assess differences in automatic annotations performed on transcribed texts, a statistical test was conducted between the groups. Subgroups were formed for dental and medical domain texts, resulting in 90 F1 score values for analysis. A Friedman test was employed to compare means, and the p-value was found to be 0.50, indicating no significant differences among the groups.

A similar comparison was performed for different entity types, also using the Friedman test, which did not reveal significant differences.

Platform Development
As a result of this research, we have developed a prototype platform [6] that integrates both components of the digital scribe system described earlier. The system then autocompletes and corrects text predictions based on context. Recording stops either manually by deactivating the microphone button or after two minutes to ensure security against open microphones.

Following recording and transcription, users can correct the transcribed text or proceed to the next step. Below the text area, there are two buttons: DELETE (to clear the text) and PROCESS (to execute the NER model for identifying medical entities such as Diseases, Body Parts, and Medications). The platform also includes a COPY button for copying identified information to the clipboard, facilitating transfer to the EHR.

The platform features a query history, enabling users to review previous queries and switch between devices during clinical care. For instance, users can record audio on a smartphone during a patient encounter and later process the data on a desktop computer, making necessary corrections and identifying entities before copying the information to the patient's clinical record.

This system promotes interoperability by allowing users to copy and paste elements into any text area of their chosen Electronic Health Record (EHR) system. Information collected from the platform can be stored for analysis and used to develop updates that better meet users' needs.

Conclusion
In this study, we have presented a unified digital scribe system for automatic recording, transcription, and annotation of clinically relevant entities from Spanish clinical audio data. We evaluated its performance in a simulated environment.

Our experimental results revealed a mean Word Error Rate (WER) among the four speakers for dental, medical, and general domains. Additionally, we found mean F1 scores for automatic entity recognition in medical and dental domain texts. Transcription errors typically involved changes in pluralization, verb forms, and pronoun usage.

The integration of these tools provides the foundation for the development of a "digital scribe" that could potentially address critical documentation-related issues such as professional burnout, doctor-patient interactions, and the time spent on documentation in clinical practice.

Future work includes exploring strategies to enhance transcription and entity detection, such as noise removal systems and pretraining speech recognition models. Usability testing will also be conducted to gather feedback from healthcare professionals [22-24].

References

1. Dalianis H (2018) Clinical Text Mining. https://library.oapen.org/handle/20.500.12657/27905.
2. Goic A (2018) Semiología Médica. https://mediterraneo.cl/medicina/4-semiologia-medica-4ed.html.
3. Stopford E, Ninan S, Spencer N (2015) How to write a discharge summary. BMJ 351: h2696.
4. Luh J, Thompson R, Lin S (2019) Clinical documentation and patient care using artificial intelligence in radiation oncology. Journal of the American College of Radiology 16: 1343-1346.
5. Quiroz JC, Laranjo L, Kocaballi AB, Berkovsky S, Rezazadegan D, et al. (2019) Challenges of developing a digital scribe to reduce clinical documentation burden. NPJ Digital Medicine https://www.nature.com/articles/s41746-019-0190-1.
6. Lai KH, Topaz M, Goss FR, Zhou L (2015) Automated misspelling detection and correction in clinical free-text records. Journal of Biomedical Informatics 55: 188-195.
7. McGoldrick M (2016) Preventing contamination of portable computers. Home healthcare now 34: 221.
8. Outomuro D, Actis A (2013) Analysis of ambulatory consultation length in medical clinics. Revista médica de Chile 141: 361-366.
9. Rabiner, LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77: 257-286.
10. Rabiner LR, Juang BH (1993) Fundamentals of Speech Recognition (Prentice Hall signal processing series). Prentice Hall, Philadelphia, PA https://www.amazon.in/Fundamentals-Speech-Recognition-Prentice-hall-Processing/dp/0130151572.
11. Quatieri TF (2001) Discrete-time Speech Signal Processing. Prentice Hall, Philadelphia, PA https://www.amazon.in/Discrete-Time-Speech-Signal-Processing-Prentice-Hall/dp/013242942X.
12. Sharma FR, Wasson S (2012) A speech recognition and synthesis tool: Assistive technology for physically disabled persons. International Journal of Computer Science and Telecommunications https://www.semanticscholar.org/paper/A-Speech-Recognition-and-Synthesis-Tool-%3A-Assistive-Sharma-Wasson/4ba27e72018532bd3e3518b800b1e648398764f8.
13. Amazon Transcribe. https://aws.amazon.com/transcribe/.
14. IBM Watson Es IA Para Empresas Más Inteligentes. https://www.ibm.com/watson.
15. Oronoz M, Gojenola K, Pérez A, de Ilarraza AD, Casillas A (2015) On the creation of a clinical gold standard corpus in Spanish: Mining adverse drug reactions. Journal of Biomedical Informatics 56: 318-332.
16. Moreno I, Boldrini E, Moreda P, Romá-Ferri MT (2017) Drugsemantics: A corpus for named entity recognition in Spanish summaries of product characteristics. Journal of Biomedical Informatics 72: 8-22.
17. Campillos Llanos L, Valverde Mateos A, Capllonch A, Moreno Sandoval A (2021) A clinical trials corpus annotated with UMLS entities to enhance the access to evidence-based medicine. BMC Medical Informatics and Decision Making https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-021-01395-z.
18. Miranda Escalada A, Gascó L, Lima López S, Farré Maduell E, Estrada D, et al. (2022) Overview of distemist at bioasq: Automatic detection and normalization of diseases from clinical texts: results, methods, evaluation and multilingual resources. https://ceur-ws.org/Vol-3180/paper-11.pdf.
19. Báez P, Villena F, Rojas M, Durán M, Dunstan J (2020) The Chilean waiting list corpus: a new resource for clinical named entity recognition in Spanish. In: Proceedings of the 3rd Clinical Natural Language Processing Workshop 291-300.
20. Báez P, Bravo-Marquez F, Dunstan J, Rojas M, Villena F (2022) Automatic extraction of nested entities in clinical referrals in Spanish. ACM Trans. Comput. Healthcare 3: 1-22.
21. Rojas M, Bravo-Marquez F, Dunstan J (2022) Simple yet powerful: An overlooked architecture for nested named entity recognition. In: Proceedings of the 29th International Conference on Computational Linguistics 2108-2117.
22. Filippidou FP, Moussiades L (2020) A benchmarking of IBM, Google, and Wit automatic speech recognition systems. Artificial Intelligence Applications and Innovations 583: 73-82.
23. Fort K (2016) Collaborative Annotation for Reliable Natural Language Processing. ISTE Ltd and John Wiley & Sons, London, England https://hal.science/hal-01324322/document.
24. Báez P, Villena F, Zúñiga K, Jones N, Fernández G, et al. (2021) Construcción de recursos de texto para la identificación automática de información clínica en narrativas no estructuradas. Revista médica de Chile 149: 1014-1022.

Copyright: ©2022 Wasim Fathima Shah. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.