Applied Computer Science 2026, vol. 22, no. 2, pp. 48–66
https://doi.org/10.35784/acs_8699

Submitted: 2025-11-08 | Revised: 2026-04-10 | Accepted: 2026-04-29

Keywords: speech recognition, controlled medications, healthcare professionals, pharmacovigilance

Luis Enrique COLMENARES-GUILLÉN [1*], *Angel Axel MÉNDEZ-MENESES* [1]
[1] Benemérita Universidad Autónoma de Puebla, Mexico, enrique.colmenares@correo.buap.mx, angel.mendezmen@alumno.buap.mx
[*] Corresponding author: enrique.colmenares@correo.buap.mx

An automatic speech recognition approach for controlled medications prescription with natural language processing

Abstract

The prescription and documentation of controlled medications require strict regulatory compliance and high transcription accuracy to prevent medication errors and ensure traceability. In many hospitals, these processes are still performed manually, increasing the risk of transcription errors, administrative delays, and non-compliance with regulatory standards, particularly for medications classified under fractions II and III of the Mexican General Health Law. Addressing this challenge requires intelligent systems capable of accurately transcribing and structuring medical prescriptions from spoken language. This study presents the design and development of an Automatic Speech Recognition (ASR) system integrated with Natural Language Processing (NLP) to support the generation and transcription of controlled medication prescriptions. The system architecture was developed following an analysis of the clinical workflow for medication requests, management, prescription, and transcription, conducted in collaboration with healthcare professionals from the hospital's Pharmacovigilance Department in Puebla, Mexico, and aligned with hospital operational standards. The methodology involved evaluating and fine-tuning three ASR models to improve transcription accuracy for medication names, dosages, and prescription instructions. NLP techniques were subsequently applied to identify and structure key prescription entities, ensuring compliance with national health regulations. Among the evaluated models, the Wav2Vec2 architecture developed by Jonatas Grosman demonstrated the best performance and was selected for implementation. Experimental results show that the optimized ASR model achieved a Word Error Rate (WER) of 6.30%, a precision of 94.72%, a recall of 91.73%, and an F1-score of 93.22%. These results demonstrate the effectiveness of the proposed approach in improving transcription accuracy while reducing false positives in prescription generation. The proposed system highlights the potential of ASR–NLP integration to enhance efficiency, accuracy, and regulatory compliance in hospital pharmacovigilance processes.

1. INTRODUCTION

The adoption of digital technologies in clinical environments has significantly transformed healthcare systems, enabling improvements in diagnosis, treatment, and patient safety across multiple medical specialties, including cardiology, telesurgery, and vital-sign monitoring (Fernández-Tapia, 2021; Jeilani & Hussein, 2025; Navarro et al., 2022). These advances have promoted the development of intelligent systems that assist healthcare professionals (HCPs) in clinical decision-making, medical documentation, and patient management. Despite these advances, several hospital processes, particularly those involving controlled medications, continue to rely heavily on manual procedures that are prone to human error.

Controlled medications are subject to strict regulatory frameworks due to their potential for abuse and dependency. In Mexico, the Ley General de Salud establishes detailed regulations for the manufacture, storage, distribution, and prescription of these medications, classifying them into different categories according to their pharmacological risk (Cámara de Diputados del H. Congreso de la Unión, 2024).

Medications belonging to fractions I, II, and III are considered highly controlled substances and require medical prescriptions with strict documentation and traceability requirements regulated by the Comisión Federal para la Protección contra Riesgos Sanitarios (2018) (COFEPRIS). Hospitals must comply with

additional operational regulations, such as the Modelo Único de Evaluación de la Calidad (MUEC), (Consejo de Salubridad General, 2023), which establishes quality, safety, and medication management standards in healthcare institutions.

Within this regulatory framework, pharmacovigilance plays a critical role in ensuring patient safety. Pharmacovigilance encompasses the detection, assessment, understanding, and prevention of adverse effects or other medication-related problems (Rai & Singh, 2024). Hospitals typically implement internal procedures, such as medication dictionaries that contain detailed drug information, including active ingredients, dosage, therapeutic indications, contraindications, and interactions. However, the preparation and transcription of prescriptions remain largely in manual processes that depend on the accuracy and diligence of healthcare personnel. These medications (Junta Internacional de Fiscalización de Estupefacientes, 2023) are classified as narcotic drugs (Moscoso Paredes & Titto Beltran, 2015) in Chapter V of Article 234 of the Ley General de Salud. COFEPRIS provides a guide for the Commercialization of controlled medications that establishes in different sections, the requirements and guidelines for marketing, dispensing, storage, safety, distribution, training, safety measures, and sanctions (Narro Robles et al., 2018).

Medication errors represent a significant global public health problem. According to the World Health Organization (WHO), approximately one in thirty patients experiences harm due to medication-related errors, with a substantial proportion being serious or potentially fatal (Hodkinson et al., 2020; World Health Organization, 2023). These errors generate an estimated global economic cost of 42 billion USD annually and can occur during multiple stages of the medication process, including prescription, transcription, dispensing, and administration (Mejía Vázquez et al., 2018; World Health Organization, 2022). In Mexico, several studies have documented high rates of prescription-related errors. For instance, an evaluation conducted at the Lic. Adolfo López Mateos Regional Hospital (ISSSTE) identified 686 medication errors, of which 84.3% were associated with prescriptions (Rey-Pineda & Estrada-Hernández, 2014). Other studies have reported a prevalence of potentially inappropriate prescriptions exceeding 70% in hospital settings (Martínez-Ruiz et al., 2023), while pediatric studies identified 776 prescription errors in 6,119 prescriptions, most of which were related to incorrect dosage (Ayuzo del Valle et al., 2021). Nationwide analyses indicate that approximately 58% of prescriptions in Mexico contain errors, ranging from therapeutic inefficiency to prolonged hospitalization (Barranco Castañeda et al., 2020).

Recent advances in Artificial Intelligence (AI), particularly in Natural Language Processing (NLP) and Automatic Speech Recognition (ASR), offer promising solutions to improve clinical documentation and reduce medication errors. NLP enables computers to interpret and analyze human language, facilitating tasks such as information extraction, semantic analysis, and automated classification of medical data (Báez et al., 2022; Villena & Dunstan, 2019; Yang et al., 2024). Data mining techniques complement NLP by identifying patterns and relationships within large datasets that support clinical decision-making (Qiao et al., 2024). In healthcare contexts, NLP has been applied to tasks such as medical text classification, electronic health record analysis, and prescription interpretation (Carchiolo et al., 2019; Masoumi et al., 2024). Similarly, speech recognition technologies have demonstrated their potential in medical environments by enabling voice-based interaction with digital systems, reducing manual data entry, and improving workflow efficiency.

Table 1 compares representative studies that apply NLP, speech recognition, or AI techniques in medical prescription analysis or healthcare documentation systems. The comparison highlights the limitations of previous work and the specific contributions of the proposed system.

Tab. 1. Comparison of related works and the proposed system

Study	Technology	Application	Dataset / context	Key results	Limitations
Carchiolo et al., 2019	NLP + Image Processing	Classification of medical prescriptions from images	Italian National Health Service prescriptions	Automated classification of prescriptions using semantic analysis	Focuses on image-based prescriptions; does not address speech transcription
Masoumi et al., 2024	BERT (NLP)	Classification of medical research texts	Biomedical literature datasets	Micro F1-score up to 85.64% for research classification	Not focused on prescriptions or medication management
Báez et al., 2022	NLP	Extraction of information from clinical texts	Medical documents	Effective extraction of clinical entities	Does not integrate speech recognition
Chala & Rebellón-Martínez, 2024	Telemedicine platforms	Remote consultation system	2,485 telemedicine consultations	96.4% resolution efficiency	Does not address prescription automation
World Health Organization, 2022; 2023; Bates et al., 1998	Computerized medical order systems	Reduction of medication errors	Hospital systems	Reduction of medication errors up to 55%	Focus on structured digital entry rather than speech transcription
Proposed System	ASR + NLP + Data Mining	Automatic transcription and structuring of controlled medication prescriptions	Hospital pharmacovigilance workflow in Mexico	WER 6.30%, Precision 94.72%, Recall 91.73%, F1-score 93.22%	Addresses regulatory compliance and speech-based prescription transcription

This study proposes an ASR–NLP framework for automated prescription generation, designed to reduce transcription errors, ensure regulatory compliance, and improve clinical efficiency.

The main scientific contributions of this research are summarized as follows:

1. Integration of ASR and NLP for controlled medication prescriptions. This work proposes a novel architecture that integrates Automatic Speech Recognition and Natural Language Processing to automate the transcription and structuring of medical prescriptions for controlled medications, addressing a critical gap in pharmacovigilance workflows.
2. Domain-adapted ASR model for medical vocabulary. The study fine-tunes a speech recognition model to improve recognition accuracy for specialized medical terminology, including medication names, dosages, and prescription instructions, which are typically difficult for general ASR systems to recognize.
3. Regulation-aware prescription processing. Unlike conventional medical transcription systems, the proposed approach incorporates regulatory constraints derived from the Mexican General Health Law and COFEPRIS guidelines, ensuring that prescriptions comply with legal requirements for controlled medications.
4. Workflow-driven system design based on real hospital processes. The system architecture is derived from an analysis of the pharmacovigilance workflow used in a hospital environment, enabling the solution to align with real-world clinical procedures rather than purely theoretical models.
5. Experimental validation with strong performance metrics. The optimized ASR model achieved a Word Error Rate (WER) of 6.30%, Precision of 94.72%, Recall of 91.73%, and an F1-score of 93.22%, demonstrating the feasibility of speech-based prescription transcription for controlled medication management.
6. Contribution to patient safety and healthcare efficiency. By reducing transcription errors and automating part of the prescription documentation process, the ASR-NLP framework improves medication safety, reduces administrative workload, and enhances the reliability of pharmacovigilance systems.

Existing works primarily focus on text-based prescription analysis or general-purpose ASR systems, lacking integration with regulatory constraints and real-time prescription generation. This gap motivates the ASR-NLP framework.

The remainder of this paper is organized as follows. Section 2 describes the methodology used for system development and model training. Section 3 presents the experimental setup and evaluation of metrics. Section 4 discusses the results obtained from the ASR and NLP models. Finally, Section 5 presents conclusions and future research directions.

2. METHODOLOGY

2.1. System architecture for controlled medication automation

The proposed framework processes speech input from healthcare professionals to generate structured prescriptions. The ASR module transcribes audio into text, while the NLP module extracts key entities, including patient information, medication details, and treatment parameters. The system incorporates validation steps to ensure compliance with controlled medication regulations.

As illustrated in Figure 1, the prescription process for controlled medications begins with the healthcare professional (HCP) reviewing the hospital's controlled medication inventory. The HCP first inspects the physical inventory and updates the corresponding records in the digital logbook. If inventory changes are detected, the HCP consults the hospital information system (MEDSYS) using the controlled medication key to retrieve usage information and identify the corresponding patient records.

Once the relevant patient information is located, the HCP initiates the prescription process. At this stage, the HCP determines whether the prescribed medication belongs to fraction I or to fractions II and III, according to the classification established by the Mexican General Health Law. If the medication corresponds to fraction I, the prescription must be completed using the official editable format provided by COFEPRIS, which requires specific regulatory documentation.

Fig. 1. General diagram of the controlled medication prescription process

If the medication belongs to fractions II or III, the HCP prepares the prescription using the institutional format. The prescription includes the following information: the date of prescription; patient identification data (name, age, and date of birth); medication information such as trade name, active ingredient, dosage (grammage), and route of administration; treatment duration and frequency; the number of packages prescribed and the quantity per package; and finally the identification of the prescribing healthcare professional, including name, professional license number, and signature.

After the prescription is completed, it is printed in two copies. The first copy is provided to the patient for dispensing at the pharmacy, while the second copy is retained by the HCP in the hospital-controlled medication for registration, ensuring traceability and regulatory compliance.

2.2. Proposed ASR–NLP prescription system

Figure 2 shows the proposed methodology for automating the prescription and transcription process using automatic speech recognition (ASR) and natural language processing (NLP).

The system starts with the HCP recording the prescription via speech input. The audio signal is processed by a previously trained ASR model, which converts the speech to text. The transcribed text is then analyzed by an NLP model to extract relevant entities from the prescription.

If the transcription is incorrect, the system prompts the HCP to repeat the prescription. If the transcription is validated, the system automatically retrieves the current date and extracts key prescription entities, including patient name, date of birth, drug name, drug dosage, treatment duration, and number of packs prescribed.

The system then determines whether the drug belongs to Fraction I or Fractions II-III. If the drug belongs to fraction I, the system generates a notification that the prescription must be registered using the official COFEPRIS editable PDF format. If the drug belongs to Fraction II or III, the system retrieves additional information from the hospital's drug dictionary, including drug presentation, trade name, and package contents. This data is used to complete the internal digital prescription format.

Finally, the system presents the generated prescription to the HCP for review before producing the required print copies.

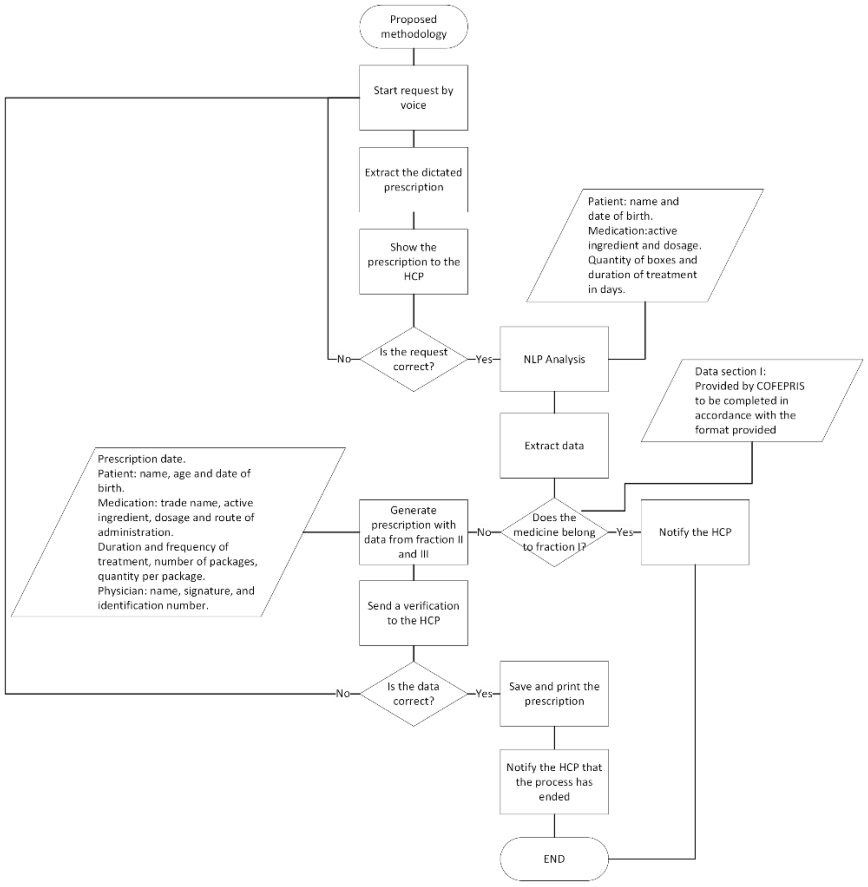

Fig. 2. The architecture of controlled medications application

2.3. NLP model training

A pre-trained NLP model was fine-tuned using Named Entity Recognition (NER) to identify domain-specific entities. The model was trained to extract key attributes, including patient name, date of birth, medication, dosage, and treatment duration. A synthetic corpus of 100,000 prescriptions was generated to train the NLP model. The dataset was split into 80% training and 20% validation.

As illustrated in Figure 3, the NLP model's training process follows a structured pipeline. The workflow begins by importing the pre-trained NLP model, after which the model configuration is adjusted by activating the NER component responsible for entity extraction. Subsequently, the dataset containing the prescription samples is loaded into the training environment.

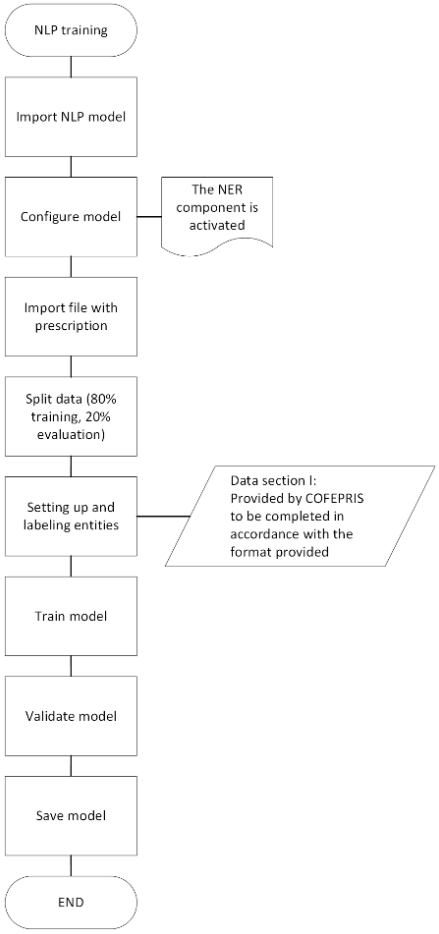

Fig. 3. Block diagram of training for the NLP model

2.4. ASR model training

The ASR model was fine-tuned using labeled audio-transcription pairs. Audio preprocessing included normalization and resampling. The model was trained to minimize transcription errors under different acoustic conditions.

The ASR training process is shown in Figure 4. First, a pre-trained speech recognition model was imported. In addition, 10,000 audio recordings (~25 hours) were collected for ASR training, accounting for variability in accents, noise conditions, and speech patterns. The dataset was split into 80% for training and 20% for validation.

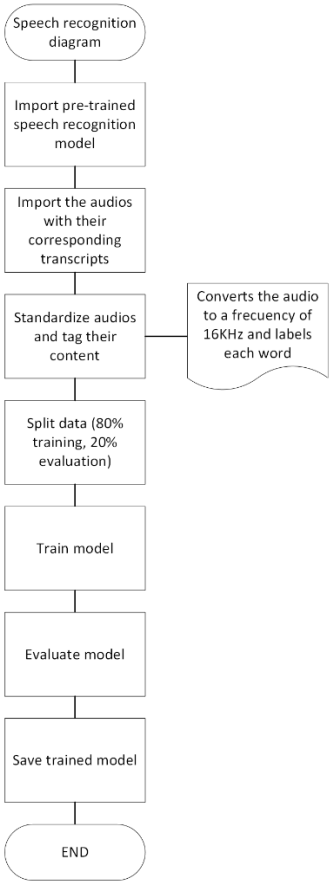

Fig. 4. Block diagram of training for ASR model

2.5. Automated prescription workflow

Figure 5 illustrates the final automation process. The HCP enters the prescription using speech. The ASR model generates a transcription, which the NLP module analyzes to identify prescription entities.

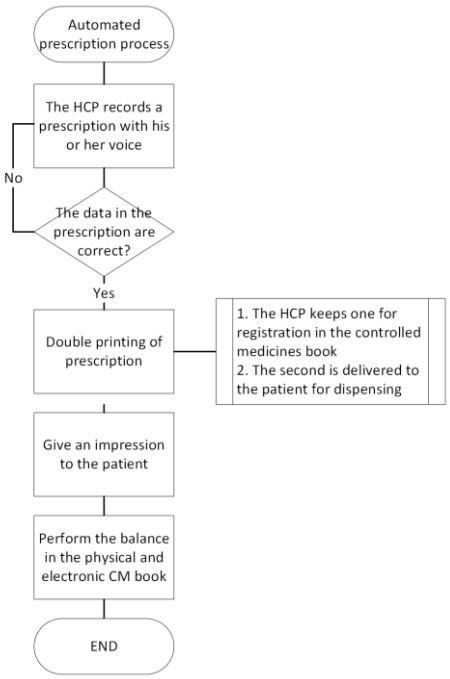

Fig. 5. Block diagram of final automation process

The HCP reviews the extracted information. If errors are found, the prescription is repeated. Once validated, the system generates a digital prescription and prints the two required copies: one for the patient and one for the hospital's drug repository.

3. EXPERIMENTAL SETUP AND EVALUATION METRICS

3.1. NLP model training for prescription entity recognition performance

The Natural Language Processing (NLP) component of the ASR–NLP framework was designed to automatically identify key entities from medical prescriptions using a Named Entity Recognition (NER) approach. This module enables the extraction of relevant clinical information from the transcribed prescriptions generated by the Automatic Speech Recognition (ASR) system.

In collaboration with healthcare professionals (HCPs), a set of 42 prescription templates was designed to represent typical prescribing patterns for controlled medications. These templates were used to generate a linguistic corpus containing structured examples of prescriptions commonly used in hospital environments.

Using these templates, a dataset of 100,000 synthetic prescription samples was generated to train the NLP model; the dataset used is shown in Table 2. In parallel, the ASR component was trained using 10,000 audio recordings containing medication names, dosage expressions, and representative prescription sentences.

Tab. 2. Dataset used for ASR and NLP training

Dataset	Description	Size
Prescription text corpus	Synthetic prescriptions generated from 42 templates	100,000 samples
Audio dataset	Voice recordings with medication names and prescriptions	10,000 audios
Training set	Used to train the NLP model	80%
Validation set	Used to evaluate model performance	20%

The prescription corpus includes annotations for the following entities: patient name, patient date of birth, medication name, medication dosage, treatment duration (days) and quantity of medication to be prescribed. These entities were manually defined and incorporated into the NER component of the NLP model.

A pre-trained NLP model with a Named Entity Recognition (NER) architecture was used as the baseline. The model was extended by introducing new domain-specific entity labels corresponding to the prescription information described above.

The NER model performs token-level classification, assigning each token in the input sentence to an entity label.

Given an input sequence of tokens:

$$X=\{x_1, x_2, \dots x_n\} \tag{1}$$

the NER model predicts the sequence of entity labels:

$$Y=y1, y2, \dots yn \tag{2}$$

where y_i corresponds to the entity label associated with token x_i.

The training objective is to minimize the cross-entropy loss function, defined as:

$$L = -\sum_{i=1}^{n} y_i log(w_i) \tag{3}$$

where: y_i - represents the true entity label.

w_i - represents the predicted probability for the entity label.

As illustrated in Figure 3, the training pipeline follows several sequential steps:

1. Importing the pre-trained NLP model.
2. Activating and configuring the NER component.
3. Loading the prescription dataset.
4. Splitting the dataset into 80% training data and 20% validation data.
5. Defining and labeling the new domain-specific entities.

6. Iteratively training the model on the annotated prescriptions.
7. Monitoring performance metrics and training losses.
8. Validating the model using the evaluation dataset.
9. Saving the optimized model for integration with the prescription system.

During training, the model iteratively processes the annotated prescriptions, updating its parameters to improve entity recognition accuracy.

The training process used the hyperparameters shown in Table 3.

Tab. 3. Hyperparameters used for NLP model training

Parameter	Value
Training samples	80,000
Validation samples	20,000
Number of epochs	30
Batch size	32
Learning rate	0.001
Optimizer	Adam
Loss function	Cross-Entropy
Entity labels	6 custom entities

Model training was monitored using the training and validation losses to ensure convergence and avoid overfitting. Table 4 shows the training losses obtained during 15 iterations of the NLP model training process. The progressive reduction in losses indicates successful model convergence during fine-tuning. The final cumulative loss was 615, corresponding to a dataset of 100,000 prescriptions and 2,387,476 tokens, yielding an average error of 2.57×10^{-4} per token. The evaluation process also measured F1 scores, which remained consistently high throughout training, demonstrating stable entity recognition performance.

Tab. 4. Loss during training

Iter	Loss	F1 Val
1	204884.6664	0.999500
2	946.7334	0.999499
3	805.2443	0.999500
4	717.2357	0.999544
5	698.7102	0.999472
6	667.9735	0.999509
7	659.5198	0.999480
8	656.4303	0.999536
9	612.2217	0.999510
10	648.7348	0.999526
11	628.4157	0.999519
12	634.8150	0.999481
13	625.9009	0.999563
14	605.6974	0.999471
15	615.3328	0.999516

After the training phase, the optimized NLP model was saved and integrated into the ASR-NLP framework, which automatically extracts relevant entities from transcribed medical prescriptions.

Figure 6 illustrates the evolution of the F1 score obtained during the training process of the proposed model. The graph shows the validation F1-score values over successive training iterations, allowing monitoring of the model's learning stability and performance.

As shown in Figure 6, the F1 score remains consistently high throughout training, fluctuating around 0.9995, indicating a strong balance between precision and recall in recognizing prescription entities. Although small fluctuations can be observed across iterations, the metric maintains a stable trend, suggesting that the model converges effectively without significant performance degradation.

These small fluctuations are expected during iterative training and are due to adjustments in model parameters as different batches of training data are processed. However, the absence of abrupt drops or unstable behavior indicates that the model maintains robust generalization performance during validation.

Overall, the results shown in Figure 6 confirm that the training procedure successfully optimizes the model parameters and achieves a consistently high F1 score, demonstrating the reliability of the ASR-NLP framework for accurate entity recognition in prescription data.

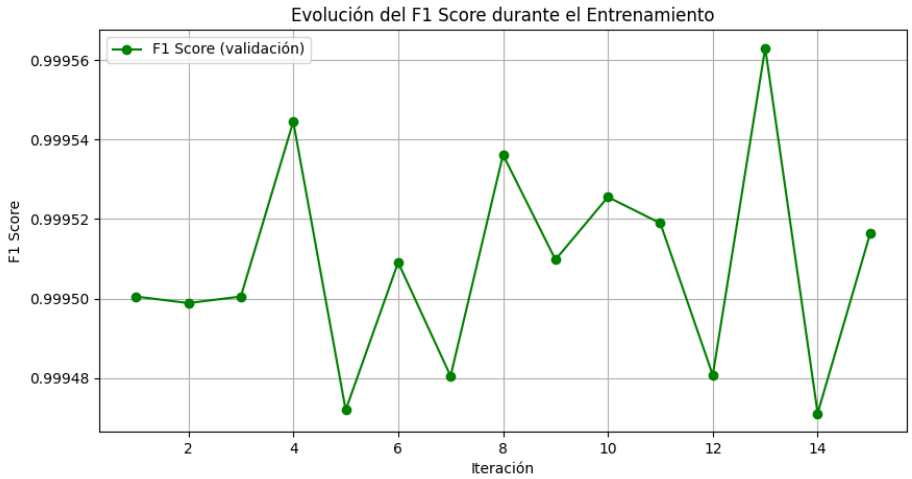

Fig. 6. F1-Score during training

3.2. NLP evaluation example

An example prescription in Spanish was evaluated: "Proporcionar 7 frascos de morfina 10 mg / 10 ml para 30 días a Esperanza Jaime Soliz nacido el 12/07/1993."

The model successfully extracted the following entities:

```
Entities:
     Nombre: Esperanza Jaime Soliz
     Medicamento: MORFINA
     Dosis: 10 MG / 10 ML
     Cajas: 7 frascos
     Días: 30 días
     Fecha_nac: 12/07/1993
```

These results demonstrate the model's ability to correctly identify structured prescription information from natural language text.

3.3. NLP model metrics

The final evaluation metrics for the NLP model are presented in Table 5.

Tab. 5. Metrics

Metrics	Micro	Macro	Pondered
Precision	0.9989	0.9989	≈0.9989
Recall	0.9998	0.9998	≈0.9998
F1-Score	0.9994	0.9994	≈0.9994

The high precision and recall values indicate reliable entity recognition with minimal prediction errors.

3.4. ASR model evaluation

The ASR model was evaluated using 24 audio recordings containing prescription commands. Table 6 summarizes the evaluation of metrics, including confidence score, Word Error Rate (WER), precision, and recall.

Tab. 6. Audio metrics

Audio	Confidence	WER	Precision	Recall
1	98.06%	0.00%	100.00%	100.00%
2	98.24%	0.00%	100.00%	100.00%
3	98.48%	0.00%	100.00%	100.00%
4	98.72%	5.56%	94.44%	94.44%
5	98.16%	0.00%	100.00%	100.00%
6	97.91%	22.73%	85.71%	94.74%
7	98.87%	4.00%	95.24%	95.24%
8	97.30%	0.00%	100.00%	100.00%
9	98.90%	0.00%	100.00%	100.00%
10	98.27%	5.00%	94.74%	94.74%
11	98.29%	8.33%	90.91%	95.24%
12	97.35%	0.00%	100.00%	100.00%
13	98.48%	8.33%	95.45%	100.00%
14	97.78%	4.17%	95.24%	95.24%
15	98.82%	0.00%	100.00%	100.00%
16	98.46%	0.00%	100.00%	100.00%
17	98.82%	4.35%	95.24%	95.24%
18	98.51%	0.00%	100.00%	100.00%
19	98.13%	36.36%	73.91%	85.00%
20	98.02%	15.00%	88.89%	84.21%
21	98.82%	0.00%	100.00%	100.00%
22	97.95%	33.33%	68.42%	72.22%
23	98.74%	4.00%	95.24%	95.24%
24	97.64%	0.00%	100.00%	100.00%

The average performance metrics were: WER 6.30%, Precision: 94.72%, Recall: 91.73%, F1-Score: 93.22%.

4. RESULTS AND DISCUSSION

The evaluation results show that the ASR-NLP framework accurately transcribes and structures recipe data.

Table 6 presents the performance metrics obtained from evaluating 24 audio samples used to test the automatic speech recognition (ASR) model. The table shows the confidence score, word error rate (WER), precision, and recall for each audio recording. Overall, the results show a high level of transcription accuracy, with most audio samples achieving WERs close to 0% and precision and recall close to 100%, indicating reliable recognition of drug names, dosages, and patient information.

However, some audio samples performed less well. The lowest-performing audio is audio 19, which achieved a WER of 36.36%, precision of 73.91%, and recall of 85.00%. The correct transcription for this audio is:

"Solicitud de medicamento clonazepam 2.5 mg / 2.5 ml 7 cajas por 7 días para José María Cornelio Valle nacido el 23/08/1977."

The transcription generated by the model was:

"solicitud de medicamento clonazepam 5 mg / d ml 7 cajas por 7 días para jose maría cornelio valle nacido el veintitres de agosto de 1977."

In this case, the model produced errors in medication dosage and the patient's date of birth, significantly increasing the WER. Such errors are particularly critical in clinical contexts because incorrect dosage information may affect patient safety.

In contrast, Audio 8 demonstrates the model's ability to produce accurate transcriptions. The correct transcription for this sample is:

"La paciente María Eugenia Mayorga nacida el 05/08/1995 durante 12 días va a tomar 6 unidades de morfina de 2.5 mg / 2.5 ml."

The transcription produced by the model was identical:

"la paciente maría eugenia mayorga nacida el 05/08/1995 durante 12 días va a tomar 6 unidades de morfina de 2.5 mg / 2.5 ml."

As shown in Table 6, this audio achieved WER = 0%, precision = 100%, and recall = 100%, demonstrating perfect transcription performance.

The differences between these two examples suggest that audio quality and pronunciation variability can influence transcription accuracy. Recordings with background noise, ambiguous pronunciation, or variations in speech rate may increase the likelihood of transcription errors. Therefore, improving the audio acquisition conditions and expanding the training dataset with more diverse speech samples could further reduce transcription errors and enhance model robustness.

Overall, the results in Table 6 indicate that the proposed ASR model achieves high transcription accuracy in most cases, while highlighting specific scenarios in which improvements in audio quality and training data diversity could further enhance system performance.

4.1. ASR model comparison

During the development of the ASR-NLP framework, several automatic speech recognition (ASR) models were evaluated to determine the most suitable architecture for transcribing medical prescriptions in Spanish. The models evaluated include Facebook/wav2vec2-large-xlsr-53-spanish, Jonatasgrosman/wav2vec2-large-xlsr-53-spanish, and OpenAI/whisper-medium.

The Wav2Vec2 architecture is a self-supervised learning model that requires relatively little labeled data to achieve competitive performance. The model first uses a convolutional neural network (CNN) to transform raw audio into latent representations at approximately 20 ms intervals. These representations are then processed by a transform encoder that captures long-range dependencies and contextual relationships in the speech signal. In addition, the model uses a Vector Quantized Variational Autoencoder (VQ-VAE) to discretize the learned representations, enabling the extraction of robust acoustic features that are less sensitive to variations in accents, noise, or speech rate (Baevski et al., 2020).

In contrast, the Whisper model uses a fully supervised learning approach trained on large multilingual datasets. Whisper follows an encoder-decoder-transformer architecture, where the encoder processes the Log-Mel spectrogram of the audio signal to extract relevant acoustic features. The decoder then generates the transcription in an autoregressive, token-by-token fashion, allowing the model to perform additional tasks such as language detection and translation (Radford et al., 2022).

Figure 7 shows the comparative evaluation of the tested ASR models using two key performance metrics: Word Error Rate (WER) and F1 score. The results show clear differences in transcription performance between the models evaluated. The Jonatasgrosman/wav2vec2-large-xlsr-53-spanish model achieved the best overall performance with a WER of 0.06098 and an F1 score of 0.84138, demonstrating superior accuracy in recognizing drug names, dosages, and prescription-related information.

In comparison, the Facebook/wav2vec2-large-xlsr-53-spanish model achieved moderate performance, with higher WER and lower F1 scores, while the Whisper-medium model performed worst in this domain. These differences can be attributed to domain adaptation, since the Jonatasgrosman implementation is specifically optimized for Spanish speech recognition.

Overall, the results shown in Figure 7 demonstrate that the Jonatasgrosman Wav2Vec2 model provides the most reliable transcription performance for the proposed prescription system, making it the most suitable candidate for integration into the automated prescription workflow.

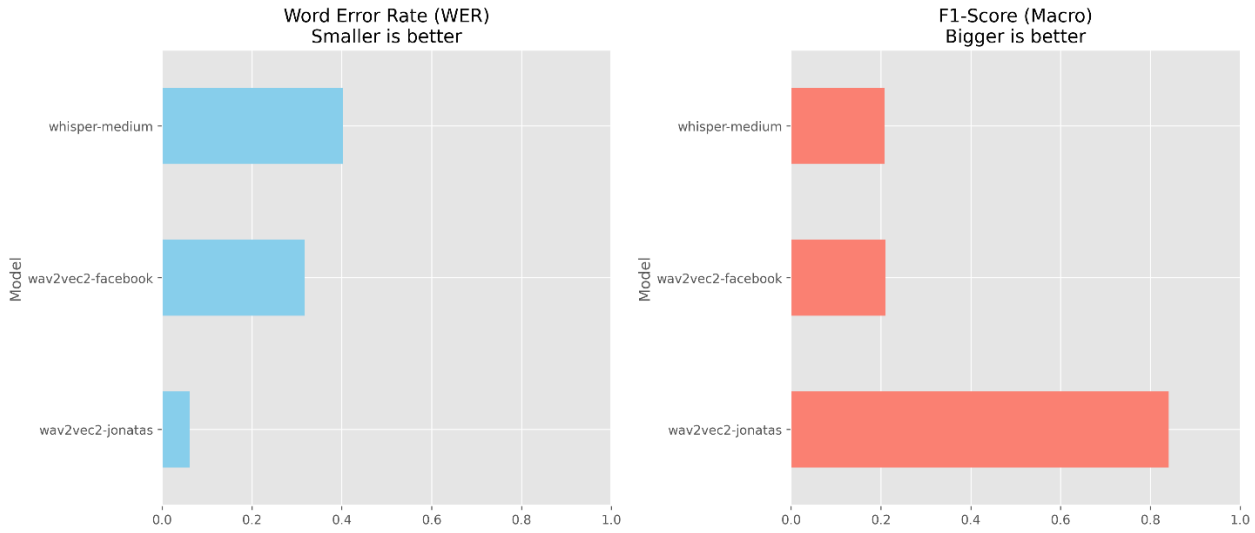

Fig. 7. ASR model comparison

4.2. User GUI

An interface with four states was designed: Record Audio, Processed Audio, Data Verification, and Generated PDF, to facilitate user interaction with the system.

Figure 8 shows the first state: Recording Audio. The buttons for recording audio are displayed. Pressing one of them starts recording from the system's default device; when finished, press the Stop Recording button. Optionally, there is a Play button to listen to the recorded or selected audio. Once the recording or playback is complete, click the Process Audio button to send it to the server.

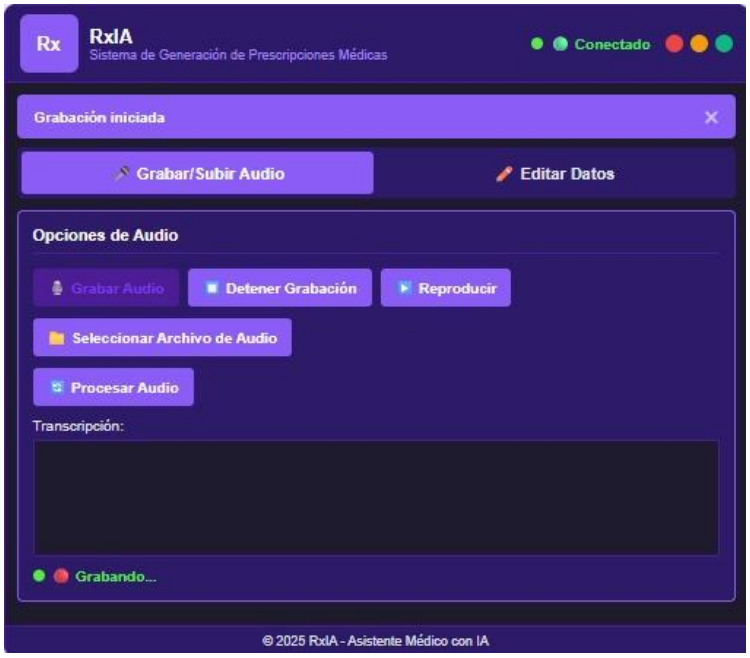

Fig. 8. Record audio

Figure 9 shows the second state: processed audio. The server receives and processes the audio. The process begins with the ASR component, which extracts the text; once complete, the text is sent to the client for display at the bottom of the screen.

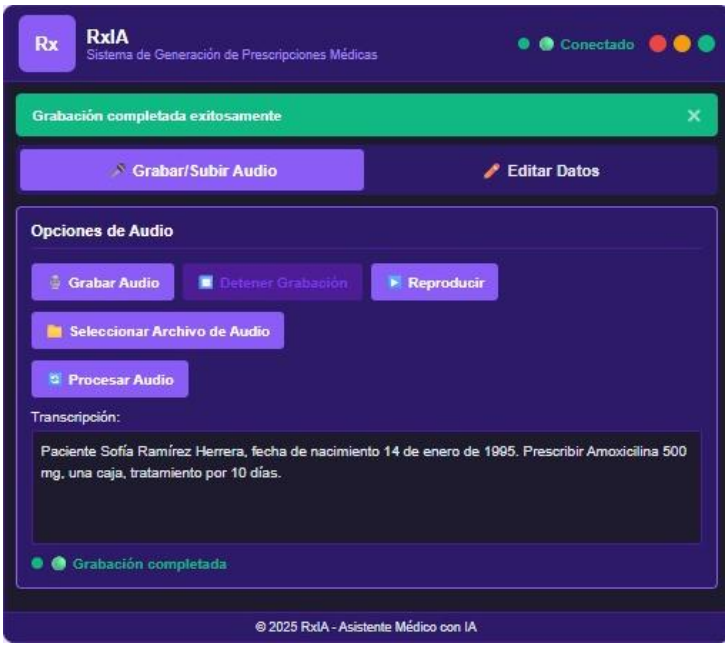

Fig. 9. Processed audio

Figure 10 shows the third state: Data Verification. The extracted text is processed by the NER component to identify existing entities. Once processing is complete, the drug information is populated using the dictionary. Finally, the information is sent to the client. In the Edit Data tab, the HCP must review the information to ensure compliance. Once the review is complete, click the Generate PDF button.

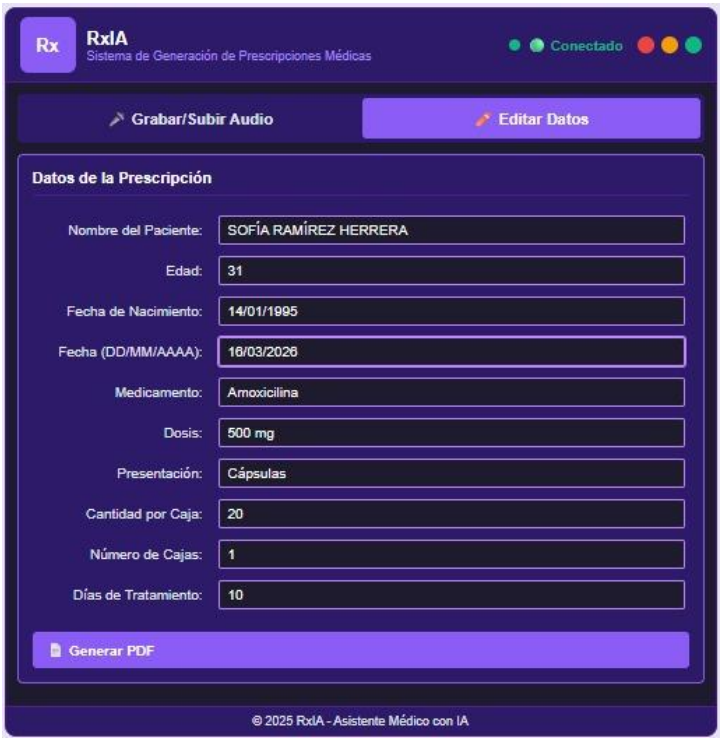

Fig. 10. Data verification

Figure 11 shows the final output: PDF generated. When the Generate PDF button is clicked, the verified information is sent to the server to generate the PDF document using the hospital's template. Finally, the generated file is sent to the client, who is notified to view and print the prescription, completing the prescription process.

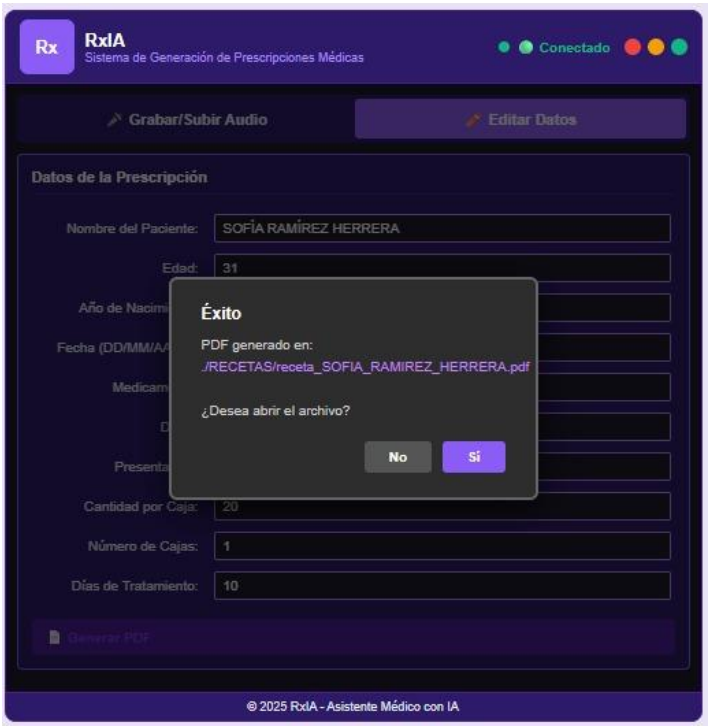

Fig. 11. PDF generated

4.3. Robustness analysis and safety mechanisms for clinical deployment

The deployment of automatic speech recognition (ASR) and natural language processing (NLP) systems in clinical environments requires careful evaluation of their robustness under uncertain, variable operating conditions. Unlike controlled experimental settings, hospital environments introduce disturbances such as background noise, accent variability, differences in speech rate, and linguistic ambiguity. These factors may degrade transcription accuracy and affect the reliability of downstream clinical decision processes.

These disturbances can be interpreted as uncertainties affecting the input signal, from a systems perspective, like perturbations studied in nonlinear and adaptive control systems. In robust control theory, system stability must be guaranteed despite bounded uncertainties and disturbances (Boulkroune et al., 2017).

Analogously, speech-based clinical systems must maintain stable performance under acoustic and linguistic variability to ensure safe and reliable operation.

The ASR training dataset incorporated variability across several dimensions to improve robustness against acoustic disturbances, including speaker characteristics, accents, speech rate, recording devices, and environmental noise. A total of approximately 10,000 audio recordings (≈25 hours of speech) were used to fine-tune the speech recognition model. These recordings contained structured prescription statements including medication names, dosage expressions, and patient information relevant to clinical prescription workflows.

Modern ASR architectures such as Wav2Vec2 and Whisper have demonstrated strong robustness due to their self-supervised or large-scale supervised training paradigms. Wav2Vec2 learns contextualized speech representations through contrastive learning and transformer-based sequence modeling.

In contrast, Whisper relies on a large multilingual dataset and a transformer encoder–decoder architecture, enabling robust speech recognition across diverse acoustic environments.

The evaluation of the selected ASR model produced the following performance metrics:

- Word Error Rate (WER): 6.30%
- Precision: 94.72%
- Recall: 91.73%
- F1-Score: 93.22%

These results indicate that the system maintains high transcription accuracy despite moderate acoustic variability.

In addition to acoustic uncertainty, clinical speech processing systems must deal with linguistic variability, including different prescription structures, dosage expressions, and drug names. To address this challenge, the NLP component of the ASR-NLP framework employs a Named Entity Recognition (NER) model trained to extract domain-specific entities relevant to prescription generation.

The system recognizes structured entities, including patient name, patient date of birth, drug name, drug dosage, treatment duration, and the quantity of the prescribed drug.

A synthetic corpus of approximately 100,000 prescription samples was generated using controlled prescription templates to train the NER component. This approach enables the model to learn consistent linguistic patterns associated with medical prescriptions while reducing ambiguity in entity extraction. Domain-specific entity recognition has been shown to significantly improve the reliability of medical NLP systems by constraining possible interpretations of clinical text (Jurafsky & Martin, 2026).

Given the safety-critical nature of prescribing systems, transcription errors must be carefully managed to prevent incorrect medical recommendations. To mitigate this risk, the ASR-NLP framework incorporates confidence-based security mechanisms derived from the ASR model's output.

Each transcription produced by the speech recognition system is associated with a confidence score representing the probability that the recognized sequence matches the spoken input. If this confidence score falls below a predefined threshold, the system can activate additional verification procedures, such as

- manual validation by the healthcare professional
- repetition of the spoken prescription
- automatic flagging of low-confidence transcriptions

Confidence-based validation serves as a supervisory control mechanism, akin to monitoring strategies used in robust adaptive systems, in which system outputs are evaluated to ensure stability and reliability under uncertainty (Boulkroune et al., 2025).

Despite the strong performance obtained during evaluation, real-world clinical environments may introduce distribution shifts, including:

- new medication names not observed during training
- rare dosage expressions
- code-switching between Spanish and English medical terminology
- unusual prescription phrasing

Handling these conditions requires additional robustness strategies such as expanded drug lexicons, multilingual training data, and stress-testing protocols. Like adaptive control systems that maintain stability under parameter variations, future versions of the system will incorporate continuous model updates and uncertainty-aware inference mechanisms to maintain reliable performance under evolving clinical conditions.

By combining acoustic variability during training, domain-specific entity recognition, and confidence-based validation mechanisms, the ASR–NLP framework aims to achieve a balance between automation efficiency and clinical safety. These design choices support the reliable generation of structured prescriptions while maintaining safeguards that prevent unsafe outputs.

The incorporation of robustness analysis and safety mechanisms is essential for the responsible deployment of systems in healthcare environments. Ensuring stable performance under uncertain conditions aligns the proposed system with established principles of robust system design and safety-critical artificial intelligence applications.

4.4. Adaptive learning and deployment considerations in dynamic clinical environments

While the fine-tuned Wav2Vec2 model demonstrates strong transcription performance, the current study assumes a relatively static linguistic and regulatory environment, which may limit long-term applicability in real-world clinical settings. In practice, pharmacological databases, prescribing standards, and legal regulations are continually evolving, particularly for controlled substances. This introduces a form of non-stationarity (distribution shift) in both acoustic and linguistic domains, where new drug names, dosage formats, and prescription structures may not be represented in the original training corpus.

This challenge is analogous to the uncertain nonlinear dynamical systems addressed in adaptive control theory, where system parameters vary over time and require continuous adjustment to maintain stability and performance. Techniques such as adaptive backstepping and neural adaptive control have demonstrated robustness to parametric uncertainty by dynamically updating model parameters based on observed deviations. Translating this paradigm to the ASR–NLP framework suggests the need for a continual learning or

incremental updating mechanism, allowing the model to incorporate new medical terminology, regulatory constraints, and prescribing patterns without catastrophic forgetting.

Future implementations should be incorporated to address this limitation:

- Incremental fine-tuning pipelines with validated prescription data,
- Dynamic lexicon expansion for emerging medications and formulations,
- Regulation-aware updates aligned with national health policies,
- and active learning strategies involving Healthcare Professionals (HCPs) for continuous validation.

Additionally, the current evaluation primarily focuses on Word Error Rate (WER) and global metrics (Precision, Recall, and F1-score). While these metrics are informative, they do not fully capture clinical correctness, particularly in safety-critical entities such as medication names, dosages, and patient identifiers. Therefore, benchmarking against state-of-the-art medical ASR systems that integrate contextual language models or domain-constrained decoding would provide a more rigorous assessment. Entity-level accuracy (e.g., NER-based evaluation) should be incorporated as a primary metric, as small transcription errors in dosage or drug names may lead to clinically significant consequences.

From a deployment perspective, several practical considerations must be addressed to transition from a research prototype to a clinically actionable system:

- Latency and real-time performance: The system must operate within clinically acceptable response times to avoid disrupting medical workflows.
- Data privacy and security: Compliance with healthcare data protection regulations (e.g., HIPAA-equivalent frameworks or national standards) is essential, particularly when processing sensitive patient information.
- Integration with Hospital Information Systems (HIS): Seamless interoperability with existing electronic health record (EHR) systems is required to ensure adoption and scalability.
- Medico-legal validation: The system must incorporate validation layers (e.g., human-in-the-loop verification, confidence thresholds) to prevent unsafe or ambiguous prescriptions from being issued automatically.

In this context, incorporating confidence-calibrated decoding and uncertainty estimation mechanisms would significantly enhance system reliability. For instance, low-confidence predictions could trigger fallback strategies such as manual verification or re-recording, aligning the system with safety-critical design principles commonly applied in clinical decision-support technologies.

Overall, embedding adaptive learning capabilities, domain-aware evaluation metrics, and robust deployment strategies would substantially strengthen the ASR-NLP framework, enable to operate reliably under real-world variability and evolving regulatory conditions.

5. CONCLUSIONS

This study presented the design and development of an ASR–NLP framework for automated transcription and prescription generation for controlled medications in a hospital environment. By analyzing real clinical workflows, key stages in the prescription process were identified and modeled, enabling the proposal of an automation strategy aligned with hospital standards and regulatory requirements.

The integration of speech recognition and named-entity recognition enabled the automatic extraction of critical prescription elements, including medication names, dosages, and patient information. The NLP model demonstrated high performance in entity recognition, while the fine-tuned ASR model achieved a WER of 6.30% and an F1-score of 93.22%, indicating reliable transcription capabilities under realistic conditions.

The ASR-NLP framework significantly reduces prescription time, from approximately 5 minutes to 15 seconds, improving operational efficiency while minimizing human error. This demonstrates potential to enhance patient safety and support healthcare professionals in high-demand environments.

A prescription generator was implemented to produce a large-scale linguistic corpus for training, supporting the development of the NLP component. This corpus included commonly used prescription structures, enabling the model to learn relevant entities, such as patient information, medication names, dosages, treatment duration, and prescription quantities. A pre-trained NER-based NLP model was adapted to recognize these domain-specific entities, enabling efficient extraction of prescription data from the transcribed text.

Despite these promising results, the system remains sensitive to acoustic variability and unseen linguistic patterns. The future work, several improvements are planned. First, the ASR and NLP models will be further

refined using larger and more specialized datasets containing medication names, dosages, and clinical expressions. Second, post-processing techniques will be incorporated to improve transcription accuracy and entity extraction. Third, the database of controlled medications will be expanded to include additional regulatory categories. Finally, this will focus on improving robustness through larger and more diverse datasets, continual learning strategies, and enhanced validation mechanisms.

Overall, the ASR-NLP framework represents a step toward intelligent, safe, and efficient clinical prescription systems.

Funding

This research was funded by the Vicerrectoria de Investigación y Estudios de Posgrado (VIEP) of the Benemérita Universidad Autónoma de Puebla (BUAP).

Acknowledgments

Aline Michelle, Pharmacobiological Chemist, and Juan Carlos López, Pharmacobiological Chemist, for their teaching and guidance on the pharmacovigilance process and hospital services.

Conflicts of interest

The authors declare that they have no conflicts of interest in this work.

REFERENCES

Ayuzo del Valle, N., González Camid, E., Villegas Macedo, F., Flores Osorio, J., & Bosques Padilla, F. (2021). Impacto del Servicio de Farmacia en la disminución de errores en la medicación en pediatría. *Revista de la OFIL, 31*(2), 161–165.

Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (2020). *wav2vec 2.0: A framework for self-supervised learning of speech representations*. *ArXiv, abs/2006.11477*. https://doi.org/10.48550/arXiv.2006.11477

Báez, P., Arancibia, A. P., Chaparro, M. I., Bucarey, T., Núñez, F., & Dunstan, J. (2022). Natural language processing for clinical text in Spanish: The case of waiting lists in Chile. *Revista Medica Clinica Las Condes, 33*(6), 576–582. https://doi.org/10.1016/j.rmclc.2022.10.002

Barranco Castañeda, G., Oropeza Cornejo, R., & Posada Galarza, M. E. (2020). Seguridad del paciente y uso de medicamentos, perspectiva del profesional farmacéutico en México enfocado en el macroproceso de la medicación. *Latin American Journal of Clinical Sciences and Medical Technology*.

Bates, D. W., Leape, L. L., Cullen, D. J., Laird, N., Petersen, L. A., Teich, J. M., Burdick, E., Hickey, M., Kleefield, S., Shea, B., Vander Vliet, M., & Seger, D. L. (1998). Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *JAMA, 280*(15), 1311–1316. https://doi.org/10.1001/jama.280.15.1311

Boulkroune, A., Boubellouta, A., Bouzeriba, A., & Zouari, F. (2025). Practical finite-time fuzzy synchronization of chaotic systems with non-integer orders: Two chattering-free approaches. *Journal of Systems Science and Systems Engineering, 34*(3), 334–359. https://doi.org/10.1007/s11518-024-5635-7

Boulkroune, A., Hamel, S., Zouari, F., Boukabou, A., & Ibeas, A. (2017). Output-feedback controller based projective lag-synchronization of uncertain chaotic systems in the presence of input nonlinearities. *Mathematical Problems in Engineering, 2017*, Article 8045803. https://doi.org/10.1155/2017/8045803

Cámara de Diputados del H. Congreso de la Unión. (2024). *Ley general de salud*.

Carchiolo, V., Longheu, A., Reitano, G., & Zagarella, L. (2019). Medical prescription classification: A NLP-based approach. In *Proceedings of the 2019 Federated Conference on Computer Science and Information Systems (FedCSIS)* (pp. 605–609). IEEE. https://doi.org/10.15439/2019F197

Chala, A. I. i Rebellón-Martínez, I. (2024). Evaluación de la experiencia de telemedicina en consulta de Cirugía de cabeza y cuello en un centro de referencia en Manizales. *Revista Colombiana de Cirugía*, *39*, 386–395. https://doi.org/10.30944/20117582.2498

Comisión Federal para la Protección contra Riesgos Sanitarios. (2018). *Guía para comercialización de medicamentos controlados en farmacias*. Gobierno de México.

Consejo de Salubridad General. (2023). *Modelo único de evaluación de la calidad: Anexo B. Criterios y estándares para hospitales*. Dirección General de Calidad y Educación en Salud. https://www.ssaver.gob.mx/ccs/wp-content/uploads/sites/35/2024/01/Anexo_B._Criterios_y_Estxndares_Hospitales._V.20-07-2023.pdf

Fernández-Tapia, J. (2021). Avances y limitaciones en las políticas públicas de e-Salud en México. *ComHumanitas: Revista Científica de Comunicación, 12*(1), 152–178. https://doi.org/10.31207/rch.v12i1.303

Hodkinson, A., Tyler, N., Ashcroft, D. M., Keers, R. N., Khan, K., Phipps, D., Abuzour, A., Bower, P., Avery, A., Campbell, S., & Panagioti, M. (2020). Preventable medication harm across health care settings: A systematic review and meta-analysis. *BMC Medicine, 18*(1), Article 313. https://doi.org/10.1186/s12916-020-01774-9

Jeilani, A., & Hussein, A. (2025). Impact of digital health technologies adoption on healthcare workers' performance and workload: Perspective with DOI and TOE models. *BMC Health Services Research, 25*(1), Article 142. https://doi.org/10.1186/s12913-025-12414-4

Junta Internacional de Fiscalización de Estupefacientes. (2023). *Lista de sustancias sicotrópicas sometidas a fiscalización internacional*.

Jurafsky, D., & Martin, J. H. (2026). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition with language models* (3rd ed. draft).

Martínez-Ruiz, M. G., Corona-Ruiz, F., Solís-Rivera, A. P., Sifuentes-Franco, S., Sánchez-López, V. A., Guevara-Martínez, S. J., & Huerta-Olvera, S. G. (2023). Potentially inappropriate prescriptions in geriatric patients hospitalized in the internal medicine department of a referral hospital in Mexico. *Gaceta Medica de Mexico, 159*(2), 150–156. https://doi.org/10.24875/GMM.22000376

Masoumi, S., Amirkhani, H., Sadeghian, N., & Shahraz, S. (2024). Natural language processing (NLP) to facilitate abstract review in medical research: The application of BioBERT to exploring the 20-year use of NLP in medical research. *Systematic Reviews, 13*(1), Article 13. https://doi.org/10.1186/s13643-024-02470-y

Mejía Vázquez, R., Delgado Cruz, F., Salgado Schoelly, H., & Kai Forzán, J. (2018). *Manejo farmacológico de las complicaciones crónicas de la diabetes mellitus (DM)*. Secretaría de Salud de la Ciudad de México.

Moscoso Paredes, A. J., & Titto Beltran, O. M. (2015). *Problemática de las drogas: Orientaciones generales*.

Navarro, E. M., Ramos Álvarez, A. N., & Soler Anguiano, F. I. (2022). A new telesurgery generation supported by 5G technology: Benefits and future trends. *Procedia Computer Science, 200*, 31–38. https://doi.org/10.1016/j.procs.2022.01.202

Qiao, H., Chen, Y., Qian, C., & Guo, Y. (2024). Clinical data mining: Challenges, opportunities, and recommendations for translational applications. *Journal of Translational Medicine, 22*(1), Article 50. https://doi.org/10.1186/s12967-024-05005-0

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). *Robust speech recognition via large-scale weak supervision*. arXiv. https://doi.org/10.48550/arXiv.2212.04356

Rai, V., & Singh, S. (2024). *A review paper on pharmacovigilance: An overview*. ResearchGate.

Rey-Pineda, E., & Estrada-Hernández, L. O. (2014). *Errores de medicación en pacientes del Hospital Regional Lic. Adolfo López Mateos del ISSSTE*.

Villena, F., & Dunstan, J. (2019). Obtención automática de palabras clave en textos clínicos: Una aplicación de procesamiento del lenguaje natural a datos masivos de sospecha diagnóstica en Chile. *Revista Médica de Chile, 147*(10).

World Health Organization. (2022). *La OMS pide a los países que actúen urgentemente para lograr la medicación sin daño*. https://www.who.int/es/news/item/16-09-2022-who-calls-for-urgent-action-by-countries-for-achieving-medication-without-harm

World Health Organization. (2023). *Seguridad del paciente*. https://www.who.int/es/news-room/events/detail/2023/09/17/default-calendar/world-patient-safety-day-2023--engaging-patients-for-patient-safety

Yang, R., Zeng, Q., You, K., Qiao, Y., Huang, L., Hsieh, C. C., Rosand, B., Goldwasser, J., Dave, A., Keenan, T., Ke, Y., Hong, C., Liu, N., Chew, E., Radev, D., Lu, Z., Xu, H., Chen, Q., & Li, I. (2024). Ascle—A Python natural language processing toolkit for medical text generation: Development and evaluation study. *Journal of Medical Internet Research, 26*, Article e60601. https://doi.org/10.2196/60601