



A CNN–BiLSTM–CTC Recognition System with Post-Processing Optimization for Handwritten English Prescription Digitization

Rijia Sultana¹

Department of Information Technology, K.J. Somaiya Institute of Technology, Sion, Mumbai, India¹

Abstract: The digitization of handwritten medical prescriptions remains a critical challenge in healthcare informatics due to the prevalence of illegible handwriting, abbreviated terminology, and inconsistent writing styles. Errors in prescription interpretation contribute significantly to adverse drug events, medication misuse, and patient safety risks. This paper presents a robust offline handwritten prescription recognition framework based on a hybrid Convolutional Neural Network (CNN), Bidirectional Long Short-Term Memory (BiLSTM), and Connectionist Temporal Classification (CTC) architecture. The proposed system combines deep visual feature extraction using ResNet34 with sequential context modeling through BiLSTM layers and alignment-free transcription using CTC decoding. To improve clinical reliability, a post-processing module integrates fuzzy string matching and pharmaceutical contextual validation through Market Basket Analysis (MBA). The framework employs a modular segmentation strategy utilizing Single Shot Detector (SSD) based word localization and line-clustering heuristics to improve recognition quality. Experimental evaluation demonstrates a Character Error Rate (CER) of 8.5%, Character Recognition Rate (CRR) of 99.20%, and Word Recognition Rate (WRR) of 93.75% for pharmaceutical terms containing 7–10 characters. The proposed system achieves high recognition accuracy while maintaining computational efficiency, requiring substantially less memory than contemporary Vision Language Models (VLMs). The framework provides a practical, deployable solution for real-time prescription digitization and integration into healthcare information systems.

Keywords: Handwritten Text Recognition; Medical Prescription Recognition; CNN–BiLSTM–CTC; Deep Learning; Optical Character Recognition; Healthcare Informatics;

I. INTRODUCTION

Healthcare systems worldwide continue to rely on handwritten prescriptions despite the increasing adoption of electronic health records. While handwritten prescriptions offer clinicians flexibility and speed at the point of care, they introduce significant challenges throughout downstream pharmaceutical workflows. Poor handwriting, inconsistent abbreviations, and complex drug nomenclature frequently result in transcription errors that may lead to incorrect medication dispensing and adverse drug events. These errors range from minor therapeutic inefficiencies to severe, and occasionally fatal, clinical complications.

Traditional Optical Character Recognition (OCR) systems have demonstrated limited effectiveness when processing handwritten medical documents due to substantial variation in writing style, cursive script, and document quality. Character level ambiguities such as confusing “mg” with “mcg,” or misreading “U” as “0” can produce clinically dangerous dosage errors. Consequently, robust handwritten prescription recognition has become an important and active research topic within medical informatics. Recent advances in deep learning have enabled substantial improvements in handwritten text recognition through the integration of convolutional and recurrent neural networks [3].

However, many high-performing approaches rely on computationally intensive transformer architectures and Vision Language Models (VLMs), which typically require specialized hardware (GPUs/TPUs) and large memory budgets. Such requirements limit their deployment in pharmacies, clinics, and healthcare facilities with constrained computational infrastructure precisely the settings in which handwritten prescriptions remain most prevalent. This study proposes a computationally efficient hybrid CNN–BiLSTM–CTC architecture specifically designed for offline handwritten English prescription recognition. The system combines deep visual feature extraction, sequential context modeling, and pharmaceutical aware post-processing to maximize recognition accuracy while preserving practical deployment feasibility on modest hardware.



The primary contributions of this research are summarized as follows:

Development of a modular handwritten prescription recognition framework that decouples text localization from sequence recognition, limiting the propagation of segmentation errors. Integration of ResNet34 visual feature extraction with BiLSTM sequential context modeling. Implementation of CTC alignment free decoding, removing the need for explicit character level segmentation. Introduction of a pharmaceutical validation layer combining fuzzy string matching and Market Basket Analysis (MBA) for co-occurrence aware drugname correction. Demonstration of high recognition accuracy (8.5% CER, 93.75% WRR) with substantially reduced computational overhead relative to VLM-based alternatives.

Research on handwritten medical text recognition has evolved considerably over the last decade. Early systems relied primarily on handcrafted features and statistical learning methods [1], but these approaches struggled with the high variability inherent in handwritten medical documents, including inconsistent stroke width, slant, and ligature formation. Bi-LSTM-based framework enhanced with Sequence Regularization / Perturbation (SRP) style data augmentation, achieving an accuracy of approximately 89.5% on a medical terminology corpus [9]. Although effective within constrained datasets, the model demonstrated limited generalization beyond the specific medical vocabularies on which it was trained. CNN-RNN-LSTM architecture combined with fuzzy search mechanisms for multilingual handwritten prescription recognition [7]. While the system improved autonomous recognition capability, its multilingual support increased computational complexity and reduced scalability to resource constrained deployments.

MEDISCAN employed a combination of conventional OCR [6], Vision Language Models, and Gemini-based processing to achieve approximately 89% recognition accuracy and 96% precision. Nevertheless, the system's dependence on VLM fallback pathways substantially increased inference time computational requirements, limiting its suitability for low-resource clinical settings. Chung and Delteil [2] presented a computationally efficient SSD and ResNet34-BiLSTM-CTC pipeline evaluated on the IAM handwriting dataset, achieving an 8.5% Character Error Rate while maintaining processing speed superior to competing transformer centric approaches. This pipeline forms the architectural basis extended in the present study toward the pharmaceutical domain.

Additional studies further demonstrated the effectiveness of CNN-BiLSTM-CTC architectures for handwritten document recognition [5] [8], though both reported performance degradation on long sequences and highly cursive text. Collectively, the literature indicates a persistent need for recognition systems that integrate domain-specific pharmaceutical knowledge directly into the decoding pipeline, rather than relying solely on generic handwriting models or computationally expensive VLM fallbacks. The present work addresses this gap through an explicit pharmaceutical validation stage combining fuzzy matching and co-occurrence-based reasoning.

Table 1 summarizes the reviewed approaches.

Table 1. KEY LITERATURE PAPERS

Study	Approach	Reported Accuracy	Key Limitation
Tabassum et al. (2021)	Bi-LSTM with SRP data augmentation	~89.5%	Limited generalization beyond constrained medical vocabulary
Pavithiran et al. (2022)	CNN-RNN-LSTM with fuzzy search, multilingual	Not directly comparable	Multilingual support increases complexity, reduces scalability
Nagasree et al. (MEDISCAN) (2026)	OCR + VLM + Gemini-based processing	~89% accuracy, 96% precision	Heavy reliance on VLM fallback increases compute cost
Chung & Delteil (2019)	SSD + ResNet34-BiLSTM-CTC (IAM dataset)	8.5% CER	Evaluated on general handwriting (IAM), not pharmaceutical text
Waly et al. (2024) Juhash et al. (2026)	CNN-BiLSTM-CTC variants	Competitive on standard sets	Degradation on long sequences and highly cursive text



II. METHODS

The proposed architecture adopts a modular design that separates text localization from sequence recognition. This design choice reduces the propagation of segmentation errors into downstream recognition tasks and improves overall system reliability, since each stage can be independently validated and tuned.

The framework consists of five primary stages, illustrated conceptually below and detailed in the following subsections:

1. Image preprocessing (grayscale conversion, normalization, denoising).
2. Text segmentation and localization (SSD-based word detection and line clustering).
3. Feature extraction (ResNet34 backbone).
4. Sequence modeling and decoding (BiLSTM + CTC with beam search).
5. Post-processing and pharmaceutical validation (fuzzy matching + Market Basket Analysis).

Image Preprocessing

Prescription images undergo a multi-stage preprocessing pipeline designed to normalize input variability before feature extraction. **Grayscale Conversion:** Input RGB images are converted into single-channel grayscale representations. The first convolutional layer of ResNet34 is adapted through weight averaging across the RGB channels, preserving compatibility with grayscale inputs while retaining stroke-level detail. **Image Normalization:** Pixel intensities are rescaled to the range [0, 1], which improves training stability and accelerates gradient convergence during optimization.

Noise Reduction and Edge Enhancement: Median filtering and edge-detection techniques are applied to suppress scanning artifacts such as paper texture, faint creases, and stamp bleed-through while preserving the fine stroke structures that are critical for accurate character discrimination.

Text Localization and Segmentation

Accurate segmentation is a prerequisite for reliable downstream recognition, since recognition errors introduced at this stage cannot generally be recovered later in the pipeline. A Single Shot MultiBox Detector (SSD) identifies word-level bounding boxes using feature maps derived from the ResNet34 backbone. Word-level detection is preferred over character- or line-level detection because it minimizes the risk of truncating or merging textual sequences, both of which are common failure modes in dense, unevenly spaced prescription text. **Greedy Line Clustering:** Detected word boxes are grouped into lines using a vertical-overlap heuristic. Additional filtering based on median height and width statistics eliminates spurious detections such as stray marks, stamps, or letterhead artifacts, thereby improving overall segmentation quality prior to recognition.

Deep Learning Architecture

ResNet34 Feature Extraction: The first component of the recognition network is a 34-layer Residual Network (ResNet34). Residual (skip) connections facilitate the training of deep architectures by mitigating the vanishing-gradient problem, enabling the network to learn richer hierarchical feature representations without degradation in training accuracy as depth increases. The network transforms input line images into dense feature maps that preserve both local stroke characteristics (curvature, thickness, slant) and global contextual information (word shape, spacing, and relative position), which together inform the sequence model in the subsequent stage.

Bidirectional LSTM Sequence Modeling

Feature maps produced by ResNet34 are sliced along the horizontal axis into a sequence of timesteps and processed by Bidirectional Long Short-Term Memory (BiLSTM) layers. BiLSTM layers analyze character dependencies in both forward and backward directions simultaneously, allowing the network to exploit contextual cues from both preceding and following characters when disambiguating a given timestep. For example, distinguishing similarly shaped characters using surrounding pharmaceutical-term context.

Connectionist Temporal Classification

Connectionist Temporal Classification (CTC) [4] enables alignment-free recognition by computing the total probability over all valid alignments between the input feature sequence and the target output label sequence, removing the need for pre-segmented, character-aligned training data. The conditional probability of a target transcription Y given an input sequence X is expressed as:

$$P(Y | X) = \sum_{A \in \mathcal{A}(X, Y)} \prod_{t=1}^T P_t(a_t | X) \quad (1)$$

where X denotes the input feature sequence, Y denotes the target transcription, $\mathcal{A}(X, Y)$ denotes the set of all valid alignment paths between X and Y (i.e., all label sequences that reduce to Y after removing blanks and repeated characters), and $P_t(a_t | X)$ is the probability of emitting symbol a_t at timestep t , as produced by the BiLSTM output layer.



A special blank token is introduced to facilitate character-boundary identification and to correctly handle repeated consecutive characters (e.g., distinguishing “ll” from a single elongated stroke). Because the number of valid alignment paths grows exponentially with sequence length, dynamic programming (the forward–backward algorithm) combined with the Log-Sum-Exp technique is used to compute Equation (1) efficiently and to ensure numerical stability during gradient-based optimization.

Post-Processing Optimization

Beam Search Decoding: Rather than greedy (best-path) decoding, the system employs beam search with a beam width of five. Although more computationally expensive than greedy decoding, beam search retains multiple candidate label sequences at each timestep and substantially improves final transcription accuracy by avoiding premature commitment to locally optimal but globally suboptimal character choices.

Fuzzy String Matching

Medical terminology frequently contains visually and orthographically similar character patterns (e.g., “amoxicillin” vs. “amoxapine”). To mitigate such recognition errors, predicted words are compared against a validated pharmaceutical reference database using Levenshtein (edit) distance. Predictions falling within an acceptable edit-distance threshold of a known drug name are corrected to that reference term, reducing residual spelling deviations and improving overall recognition reliability.

Market Basket Analysis

Market Basket Analysis (MBA) provides contextual pharmaceutical validation by exploiting drug co-occurrence statistics mined from historical prescription data. Frequently co-prescribed combinations — such as Amoxicillin and Clavulanate — are assigned higher confidence scores when recognition ambiguity exists between multiple candidate drug names on the same prescription. This co-occurrence-aware reasoning mechanism allows the system to resolve otherwise ambiguous predictions using domain-specific prescribing patterns, meaningfully improving final prediction quality beyond what character-level correction alone can achieve.

III. RESULTS

Overview of Findings

The model was trained using the Adam optimizer over 50 training epochs, with data partitioned into a 90:10 training validation split. Pharmaceutical term evaluation focused on words containing 7–10 characters, representative of common drug-name lengths.

Performance Metrics

Evaluation employed four standard handwriting-recognition metrics:

Character Error Rate (CER): The edit-distance-normalized proportion of character-level insertions, deletions, and substitutions.

Word Error Rate (WER) : The proportion of predicted words that do not exactly match the ground truth.

Character Recognition Rate (CRR): The proportion of correctly recognized characters.

Word Recognition Rate (WRR): The proportion of correctly recognized complete words.

Table 2. Recognition performance: baseline vs. proposed system

Metric	Baseline System	Proposed System
Character Error Rate (CER)	28.3%	8.5%
Character Recognition Rate (CRR)	—	99.20%
Word Recognition Rate (WRR)	~60%	93.75%
Precision	82%	96%

Note: Baseline figures correspond to a conventional OCR pipeline without SSD-based segmentation or pharmaceutical-aware post-processing; CRR was not separately reported for the baseline configuration.

IV. DISCUSSION

The most significant improvement was observed in full-page prescription processing, where the introduction of line clustering heuristics reduced CER from 28.3% to approximately 9.9%. The pharmaceutical linguistic denoiser (fuzzy



matching combined with Market Basket Analysis) contributed a further 1.4-percentage point reduction in error rate, yielding the final reported CER of 8.5%. The resulting WRR of 93.75% demonstrates the effectiveness of the integrated post-processing framework specifically for pharmaceutical terminology recognition, where domain aware correction provides a measurable advantage over generic handwriting recognition pipelines.

The proposed CNN–BiLSTM–CTC framework successfully addresses key limitations of conventional OCR systems and offers a practical alternative to resource-intensive transformer-based models for handwritten prescription digitization. One of the primary strengths of the system lies in its balanced, modular design. The CNN component effectively captures spatial handwriting features, while the BiLSTM layers model contextual dependencies between characters across the sequence. The CTC decoder eliminates the need for precise character-level segmentation during training, making the framework robust to the highly variable stroke patterns typical of clinical handwriting. The pharmaceutical-aware post-processing stage further distinguishes the proposed approach from generic handwriting recognition systems. By integrating fuzzy matching and co-occurrence-based validation, the framework incorporates domain knowledge directly into the recognition pipeline rather than treating recognition as a purely generic sequence-labelling problem. Compared with Vision Language Models, the proposed architecture demonstrates substantially lower memory requirements and inference latency while maintaining clinically acceptable recognition performance. These characteristics make the system well suited for deployment in pharmacies, hospitals, and low-resource healthcare environments where GPU-backed VLM inference is not economically or operationally feasible. A limitation of the present evaluation is that reported metrics are drawn from a single experimental configuration (90:10 split, 50 epochs) without cross-validation or an explicit held-out test set drawn from multiple institutions; broader validation across diverse handwriting styles, prescription formats, and clinical settings is necessary before large-scale deployment.

V. CONCLUSION

This paper presented a robust deep learning framework for offline handwritten English prescription digitization. The proposed system combines ResNet34 feature extraction, BiLSTM sequence modeling, CTC-based transcription, and pharmaceutical aware post-processing to achieve high recognition accuracy while maintaining computational efficiency suitable for real world clinical deployment. Experimental evaluation demonstrated a Character Error Rate of 8.5%, Character Recognition Rate of 99.20%, and Word Recognition Rate of 93.75%, validating the effectiveness of the architecture for healthcare applications. The modular design and reduced computational requirements make the framework a viable candidate for deployment in clinical and pharmacy environments with constrained infrastructure. Future research directions include; Expansion of pharmaceutical lexicons to better handle out-of-vocabulary and newly approved drug names. Development of mobile and cloud-based deployment APIs for point-of-care digitization. Integration with hospital information systems and electronic health records for seamless clinical workflows. Realtime synchronization with pharmacy inventory databases to flag stock or substitution issues at recognition time. Exploration of hybrid transformer CNN architectures to further improve accuracy while monitoring the resulting computational trade-offs.

REFERENCES

- [1]. Ahmad, I., Fink, G. A., & Mahmoud, S. A. (2021). Improvements in sub-character HMM model based Arabic text recognition.
- [2]. Chung, J., & Delteil, T. (2019). A computationally efficient pipeline approach to full page offline handwritten text recognition.
- [3]. Fasha, M., Hammo, B., Obeid, N., & AlWidian, J. (2020). A hybrid deep learning model for Arabic text recognition.
- [4]. Graves, A., & Hannun, A. (2017). Sequence modeling with CTC.
- [5]. Juhash, R. A., Abusonoun, A. S., & Al-Desouky, S. A. (2026). Hybrid CNN–BiLSTM–CTC framework for Arabic handwritten recognition.
- [6]. Nagasree, P., Pricilla, P., & Shreyas, K. M. (2026). MEDISCAN: AI-powered prescription scanner with real-time medicine availability.
- [7]. Pavithiran, G., et al. (2022). Doctor's handwritten prescription recognition system using deep learning.
- [8]. Waly, A., et al. (2024). Arabic handwritten document OCR solution with adaptive scale fusion detection.
- [9]. Tabassum, S., et al. (2021) Bi-LSTM based handwritten medical text recognition with SRP data augmentation.