A Critical Review of Trustworthy Handwritten Prescription Intelligence: Deep Recognition, Clinical Risk and Privacy-Aware Provenance

Kunal Kumar Mandal
Assistant Professor
Department of Computer Science
Mankar College, West Bengal, India

Abhoy Chand Mondal, PhD
Professor
Department of Computer Science,
The University of Burdwan, West Bengal, India

ABSTRACT
Handwritten prescriptions are difficult to digitize safely because visual transcription, medicine identification and clinical interpretation fail in different ways. This critical narrative review integrates research on convolutional recurrent recognition, connectionist temporal classification, attention-based decoding, multilingual prescription understanding, explainability and privacy-preserving record management. Rather than ranking incomparable accuracy figures, the analysis separates image modality, target field, evaluation design and clinical consequence. It develops a risk-aware, human-supervised reference workflow in which uncertain medication instructions are reviewed before a verified record is authenticated or shared. The synthesis argues that character and word error rates must be complemented by medication-field accuracy, critical-error detection, calibration, abstention and reviewer workload. Blockchain-based provenance may protect an approved record against undetected alteration, but cannot establish the correctness of the initial transcription. Priorities include representative Indian prescription datasets, institution-independent testing, clinically grounded explanations and proportionate security controls. This paper offers a conceptual synthesis, not new experimental or clinical validation.

General Terms
Pattern Recognition, Medical Document Intelligence, Clinical Safety, Information Security

Keywords
Handwritten prescription; critical review; OCR; CNN–BiLSTM; CTC; transformer; clinical NLP; explainable AI; blockchain

1. INTRODUCTION
A handwritten prescription is not merely a collection of characters. It encodes an intended treatment in which a single altered digit, unit or medicine name can have disproportionate consequences. Unconstrained handwriting, non-standard abbreviations, camera artefacts and multilingual content therefore challenge both visual transcription and clinical interpretation. The relevant technical question is not only whether text is recognized, but whether an uncertain system can avoid presenting a hazardous guess as a verified instruction.

The literature has evolved from isolated medicine-name recognition [1] to CNN–recurrent pipelines [2–5], attention-based approaches [8] and document-level extraction [9]. Two earlier works by the present research group offer complementary perspectives: a comparative review of deep-learning approaches [13] and the PharmaChain-Secure (PCS) proposal for cryptographically governed prescription records [14]. The present review differs by treating recognition, clinical verification and information integrity as distinct, jointly necessary stages and by explicitly analysing evidence comparability and deployment failure modes.

2. REVIEW SCOPE AND QUESTIONS
This paper is a structured critical narrative review, not a systematic review or meta-analysis. Its starting corpus comprises the two author-supplied publications [13,14] and their cited prescription-recognition studies, complemented by established sequence-recognition work [6,7]. Studies are organized by input modality, learning architecture, output granularity, evaluation metric and deployment setting. Where primary-study metrics could not be independently checked, numerical rankings are deliberately avoided. Preprints are identified as such. The synthesis addresses: (RQ1) which methods fit different prescription-recognition subtasks; (RQ2) which errors propagate into medication instructions; (RQ3) how uncertainty, human verification and provenance should interact; and (RQ4) what evidence is required for real-world adoption.

3. PRESCRIPTION COMPLEXITY AND FAILURE MODES
3.1 Visual and linguistic variability
Images vary in focus, illumination, rotation, paper texture and ink quality. Physicians may write connected cursive, overwrite characters, mix scripts or place instructions outside expected fields. Medicine names are often rare in general handwriting corpora; abbreviations and shorthand are context-dependent. Segmentation errors can propagate into word recognition and subsequent extraction.

3.2 Clinical ambiguity and asymmetric risk
Errors are not equally harmful. A minor misspelling in a heading differs from a substituted drug, wrong strength or altered frequency. Similar-looking medicine names, omitted units and unclear decimal separators are particularly important. Language models may produce plausible but unsupported corrections, making confidence calibration and explicit abstention necessary. A machine-readable output must not be mistaken for a clinically validated order.

3.3 Data and governance constraints
Prescription images can reveal names, addresses, diagnoses and other sensitive data. Consent, de-identification, access controls and institutional approvals are prerequisites for dataset release or model training. Regional variation, especially across Indian institutions and scripts, also limits transfer from public English handwriting benchmarks.

4. RECOGNITION ARCHITECTURES AND EVIDENCE

4.1 Image acquisition and localization

Scanned documents and mobile photographs differ in skew, illumination, blur and compression. Grayscale conversion, deskewing, denoising and local contrast enhancement can improve legibility, but aggressive thresholding can erase faint strokes or decimal points. Line- or region-level detection should preserve coordinates so a reviewer can return to the source image. Layout detection, including object-detection approaches, is useful when medicine names and dosing fields occupy irregular regions; however, successful localization does not imply correct text recognition [13].

4.2 CNN, recurrent and CTC recognition

Convolutional encoders learn visual stroke features, while bidirectional LSTMs model character sequences. Connectionist temporal classification (CTC) permits alignment-free training of variable-length transcriptions [6]. Prescription studies using CNN–BiLSTM or CRNN architectures [2–5] illustrate the appeal of this modular design, but their results reflect different corpora, handwriting modalities and reporting conventions. Online pen trajectories [4] should not be equated with photographed paper prescriptions. Character error rate (CER) and word error rate (WER) remain essential, yet neither directly measures correct dose extraction.

4.3 Attention, multilingual models and clinical parsing

Attention-based handwriting recognition can model long-range image–text dependencies [7], while recent prescription-focused work explores multi-head attention and multilingual decoding [8,10]. Such approaches may recover difficult sequences but can also output plausible text unsupported by visible strokes. A downstream clinical NLP stage should therefore extract medicine name, formulation, strength, dose, route, frequency and duration while retaining the original OCR candidates. Lexicon-based correction must not silently replace one valid drug with another. Language-specific abbreviations and mixed scripts require dedicated evaluation rather than extrapolation from English benchmarks.

4.4 Comparative synthesis

Table 1 is an evidence map, not a league table. Accuracy figures from incompatible datasets, splits or task definitions cannot establish an overall winning architecture. The earlier comparative review [13] reports several high percentages, but even within its summary, some study-level values differ across sections; consequently, this synthesis emphasizes reproducible task definitions and error categories instead of a cross-study accuracy ranking.

4.5 Cross-study comparability

Published systems address distinct tasks: scanned paper prescriptions, online pen trajectories, isolated medicine words, multilingual decoding and document-level extraction. Their accuracy figures should not be pooled without common data, splits and target definitions. Table 1 maps representative evidence to the most consequential interpretation limits.

Table 1.Selected studies and limits of cross-study comparison.

Study	Technical focus	Interpretation limit
Dhande&Kharat [1]	SVM / cursive words	Limited medicine-word task
Fajardo et al. [2]	Deep cursive OCR	External capture shift
Jain et al. [3]	CNN–BiLSTM	Dose-field safety untested
Tabassum et al. [4]	Online medical words	Not photographed writing
Shahade et al. [5]	CNN prescription OCR	Cross-site comparability
Pavithiran et al. [10]	Multilingual decoding	Preprint validation
Ali et al. [8]	Multi-head attention	Medicine-name extraction task; external validation needed
Reddy et al. [9]	Multimodal XAI	Clinical explanation testing

5. CLINICAL ERROR PROPAGATION

A prescription pipeline can fail at capture (missing edge), detection (omitted line), recognition (wrong character), normalization (wrong brand-name substitution), entity linking (incorrect drug concept), or presentation (uncertain output displayed as fact). These stages are causally connected: a missing decimal point can survive perfect downstream formatting. Error severity is asymmetric; a minor spelling error in a heading differs from a drug-strength error. For this reason, review datasets should label not only ground-truth text but also critical medication fields and ambiguity requiring prescriber consultation.

6. HUMAN SUPERVISED VERIFICATION WORKFLOW

Figure 1 summarizes a conceptual workflow synthesized from the recognition literature [1–13] and the separation of provenance from transcription argued in PCS [14]. It is not an implemented or clinically validated system. The design maintains traceability between each extracted field and its image region and routes unresolved critical fields to an authorized reviewer.

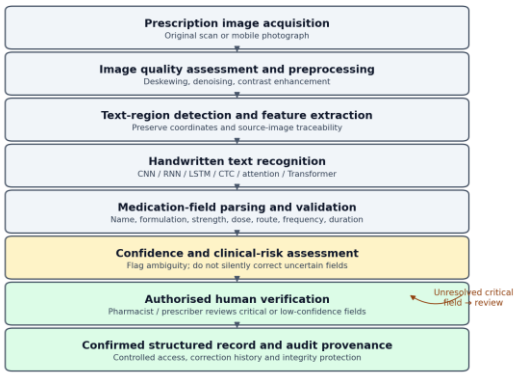


Figure 1. Conceptual human-supervised workflow for handwritten medical prescription recognition and validation.

6.1 Verification algorithm (conceptual)

Input: prescription image I, authorized medicine lexicon L, locally validated confidence policy T. (1) Reject or reacquire unreadable images. (2) Detect regions and retain bounding boxes. (3) Produce OCR candidate strings and calibrated confidence estimates. (4) Parse medication fields without overwriting source strings. (5) Compare candidates with L; flag ambiguous names, units and contradictions. (6) If a critical field is missing, inconsistent or below T, abstain and request

pharmacist/prescriber verification. (7) Otherwise create a provisional record for confirmation. (8) Preserve correction provenance and access-controlled audit metadata. Output: confirmed structured record or unresolved review case. Thresholds must be established on representative validation data; this algorithm does not authorize automated dispensing.

6.2 Conceptual decision sequence

Image capture and quality gate → text-region detection with coordinates → OCR candidates and confidence → medication-field parsing → lexicon and consistency checks → pharmacist/prescriber review when critical information is ambiguous → confirmed structured record → controlled provenance and audit. This sequence is a design recommendation, not an implemented or clinically approved dispensing system.

7. PRIVACY, BLOCKCHAIN AND PROVENANCE

The authors' published PharmaChain-Secure (PCS) chapter [14] proposes a permissioned network with doctor, patient, pharmacist and regulator roles; smart contracts, decentralized identifiers, zero-knowledge proofs (ZKPs), homomorphic encryption (HE) and off-chain storage are discussed as privacy and integrity mechanisms. Its five-node simulation reports approximately 1.2 s median issuance latency, 0.8 s transfer latency and 50 transactions/s under its stated setup [14]. These are previously published simulated system metrics, not recognition results or independently replicated clinical deployment evidence.

For prescription OCR, a hash or signature can authenticate a stored record and document its later modification; neither establishes whether the original handwriting was read correctly. Patient data should not be placed unencrypted on a permanent public ledger. Permissioned governance, encrypted off-chain storage, selective disclosure, revocation and key recovery warrant explicit threat modelling. ZKPs may prove eligibility without revealing an underlying identifier, whereas HE is useful only where an actual encrypted computation justifies its computational and implementation cost. Standard authenticated databases with digital signatures may be preferable when decentralization adds little value.

7.1 Security is not semantic correctness

A cryptographic signature can bind a verified record to an issuer and reveal subsequent tampering. It cannot distinguish a faithful transcription from a mistaken but properly signed one. Permissioned ledgers, off-chain encryption and selective disclosure should therefore be evaluated against simpler signed databases under explicit threat models; decentralization is not automatically warranted.

8. EVALUATION FRAMEWORK

Recognition research often reports character error rate (CER) or word error rate (WER). For edit distance d and reference length N, CER = d(character sequences)/N and WER = d(word sequences)/N. These metrics are necessary but not sufficient for prescription safety. Table 3 recommends complementary measures and explicit review outcomes.

At minimum, evaluation should use institution- and writer-separated splits, report sample sizes and confidence intervals, disclose exclusion rules, and examine low-quality images and multilingual cases. Human reviewers should independently adjudicate a subset of errors. Real-world deployment additionally needs prospective workflow testing, cybersecurity review and clear accountability for the final prescription decision.

8.1 Proposed reporting matrix

Table 3. Proposed multi-level evaluation criteria.

Evaluation layer	Recommended outcomes
Transcription	CER, WER; per-writer and per-site
Medicine identity	Exact match, top-k recall, look-alike confusion
Instruction fields	Strength, dose, route, frequency, duration F1
Uncertainty	Calibration, abstention and false acceptance
Clinical workflow	Critical-field errors, reviewer time, near misses
Governance	Access failures, latency, recovery, audit integrity

8.2 Results of the Critical Synthesis

The critical synthesis identifies three recurring patterns across the reviewed studies. First, recognition architectures have progressed from task-specific classifiers and cursive-word recognition towards CNN–recurrent pipelines, CTC-based sequence learning, attention mechanisms and multilingual or document-level approaches [1–5, 8–12]. These developments expand the range of handwriting patterns and prescription content that can be processed; however, the studies address different input modalities and output tasks. In particular, online pen-trajectory recognition cannot be treated as equivalent to recognition from photographed or scanned paper prescriptions.

Second, the evidence does not support a reliable ranking of the reviewed architectures by a single accuracy value. Differences in datasets, writer populations, image acquisition conditions, language, annotation policy, train–test split and target granularity limit direct comparison. Character- or word-level scores may also conceal errors in clinically critical fields. Accordingly, the evidence map in Table 1 is interpreted qualitatively, and the evaluation criteria in Table 3 are proposed to support more clinically meaningful future comparisons.

Third, the reviewed work indicates that recognition performance alone is insufficient for safe prescription digitisation. Errors can originate during image capture, text-region detection, transcription, terminology normalisation, field extraction or final presentation. A system may produce fluent text while still misidentifying a medicine, strength or dosing instruction. The most defensible design therefore preserves the original image and candidate text, reports uncertainty, avoids silent substitution, and routes unresolved critical fields to an authorised human reviewer.

Taken together, the results of this review support a layered interpretation of system quality: visual transcription, medicine-entity identification, instruction-field extraction, robustness under domain shift, uncertainty handling, human-review effectiveness and information governance should be assessed separately. This is a synthesis of the cited literature and a proposed evaluation perspective; it is not the result of a new experiment conducted by the authors.

8.3 Evaluation Across Datasets and Real-World Scenarios

A stronger evaluation programme should include datasets collected from multiple healthcare institutions and should

represent a broad range of prescribers, handwriting styles, prescription layouts, medical specialties and acquisition devices. Where ethically and legally permissible, dataset documentation should state the source institution, collection period, image modality, language or script, annotation procedure, consent or other lawful basis, de-identification method, and known limitations. The number of prescriptions, writers and institutions should be reported separately so that dataset scale and diversity are transparent.

Evaluation should distinguish scanned or photographed paper prescriptions from online handwriting captured through a digital pen. It should also include challenging image conditions such as blur, low resolution, skew, uneven illumination, faded ink, compression artefacts, overlapping strokes and partially obscured text. For multilingual settings, testing should cover mixed-language and mixed-script content, regional abbreviations and medicine names that may be uncommon in general-purpose handwriting corpora. These scenarios are particularly relevant to Indian healthcare settings, where prescription practices and language use can vary across regions and institutions.

To assess generalisation, training, validation and test partitions should be separated at writer and, where feasible, institution level. An external test set from an institution not represented during training is preferable to relying exclusively on random image-level splits. Results should report sample counts, confidence intervals, exclusion criteria and error breakdowns by field and scenario. A human-adjudicated subset should be used to resolve ambiguous ground truth and to estimate the frequency and severity of clinically important errors.

Because the present article is a critical narrative review and does not introduce a newly collected dataset or run a new recognition experiment, it does not claim comparative numerical performance across these scenarios. Rather, the scenarios above define a reproducible evaluation agenda for future studies. Any subsequent quantitative comparison should use the same benchmark, annotation rules and split protocol for all compared methods, and should report medicine-name and instruction-field performance in addition to CER and WER.

9. OPEN RESEARCH GAPS

9.1 Representative Indian datasets

A major bottleneck is access to ethically curated, multi-institutional prescriptions representing diverse writers, pharmacies, layouts, brands and regional scripts. Useful dataset documentation would include collection context, annotation guidelines, inter-annotator agreement, consent basis and limitations. Synthetic samples can supplement training but should not replace testing on real clinical handwriting.

9.2 Robustness and domain shift

A model trained on one hospital may degrade on another scanner, language, specialty or prescribing style. External evaluation, out-of-distribution detection and continual monitoring are more informative than a single random-split score. Error analyses should separate poor image capture from recognition, entity extraction and clinical validation failures.

9.3 Privacy-preserving learning

Federated learning can reduce the centralisation of raw prescription data, but it does not automatically prevent membership inference, gradient leakage or institutional bias. Secure aggregation, access control and privacy accounting should be evaluated against centralised baselines, with communication, computational and maintenance costs reported.

9.4 Clinically meaningful trust

The central research question is not whether an OCR model can produce fluent text, but whether a clinician can safely detect and correct its errors. Future work should test calibrated uncertainty, explanation usefulness, pharmacist workload, rare medication names and human–AI disagreement. An abstention-capable system may be preferable to one that maximizes apparent completeness.

10. DISCUSSION

The combined literature suggests that prescription digitization is best understood as a sequence of evidence-preserving decisions, not an end-to-end text-generation task. CNN–recurrent and transformer systems address different visual challenges; downstream terminology models can improve structured extraction but can also amplify confident mistakes. The strongest next studies will compare architectures on identical institution-disjoint datasets and report how often a reviewer catches an unsafe suggestion. The PCS proposal [14] contributes a governance perspective, yet the ordering of operations is crucial: clinical content must be confirmed before its provenance is treated as authoritative. This distinction is the central integration principle of the present synthesis.

11. CONCLUSION

Handwritten prescription recognition has progressed beyond isolated character classification towards document-level understanding, yet the evidence reviewed in this paper indicates that clinical readiness remains insufficiently established. Prescription images contain variable handwriting, non-standard abbreviations, multilingual content and clinically consequential fields. Consequently, a system that achieves plausible text transcription cannot, by that fact alone, be considered a reliable source of medication instructions. Recognition, clinical interpretation, human verification and information provenance must be treated as distinct but connected stages.

The reviewed literature demonstrates the relevance of CNN–recurrent architectures, CTC-based sequence learning, attention mechanisms, multilingual decoding and document-level extraction. Nevertheless, differences in datasets, modalities, task definitions and evaluation protocols prevent a defensible overall ranking of the reported approaches. Future studies should therefore establish ethically curated, multi-institutional prescription datasets with diverse writers, regional scripts, prescription layouts and acquisition conditions. Dataset documentation, consistent annotation guidelines, writer- and institution-independent splits, external validation and transparent reporting of sample sizes and confidence intervals are essential for reproducible evaluation.

Future research should also move beyond aggregate CER and WER by measuring medicine-name identification, formulation and strength recognition, dose, route, frequency and duration extraction, clinically critical error rates, calibration, false acceptance and abstention. Robustness should be tested under blur, low resolution, faded ink, skew, mixed scripts and domain shift. Systems should preserve source-image traceability and candidate alternatives, and should refer uncertain or high-risk fields to an authorised pharmacist or prescriber rather than silently replacing ambiguous text. Prospective workflow studies should examine reviewer workload, correction rates, near misses and accountability before any clinical deployment is considered.

Explainable artificial intelligence may help reviewers understand model outputs and identify unreliable predictions, while privacy-preserving learning and proportionate security controls may

support responsible use of sensitive prescription data. Federated learning, secure aggregation, access control and privacy accounting require evaluation against centralised alternatives. Similarly, blockchain or other provenance mechanisms may help record authorisation and subsequent changes, but they cannot establish the semantic correctness of the original transcription. Cryptographic integrity must therefore follow, and never substitute for, clinical verification.

In conclusion, trustworthy handwritten prescription intelligence requires an integrated approach combining robust multilingual recognition, clinically meaningful evaluation, calibrated uncertainty, human supervision, explainability and privacy-aware provenance. Progress towards these goals may support safer and more accountable healthcare documentation; however, claims of clinical effectiveness must ultimately be established through independent, representative and prospective validation.

12. REFERENCES

[1] P. S. Dhande and R. Kharat, "Character Recognition for Cursive English Handwriting to Recognize Medicine Name from Doctor's Prescription," ICCUBEA, 2017.

[2] L. J. Fajardo et al., "Doctor's Cursive Handwriting Recognition System Using Deep Learning," 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), 2019, doi: 10.1109/HNICEM48295.2019.9073521.

[3] T. Jain, R. Sharma and R. Malhotra, "Handwriting Recognition for Medical Prescriptions using a CNN-Bi-LSTM Model," 2021 6th International Conference for Convergence in Technology (I2CT), 2021, doi: 10.1109/I2CT51068.2021.9418153.

[4] S. Tabassum et al., "An online cursive handwritten medical words recognition system for busy doctors in developing countries for ensuring efficient healthcare service delivery," Scientific Reports, vol. 12, art. 3601, 2022, doi: 10.1038/s41598-022-07571-z.

[5] M. Shahade et al., "Convolutional Neural Networks for Handwritten Text Recognition of Medical Prescription," Journal of Digital Information Management, vol. 21, no. 4, pp. 117–124, 2023, doi: 10.6025/jdim/2023/21/4/117-124.

[6] A. Graves et al., "Connectionist Temporal Classification: LabellingUnsegmented Sequence Data with Recurrent Neural Networks," ICML, 2006.

[7] L. Kang et al., "Pay Attention to What You Read: Non-recurrent Handwritten Text-Line Recognition," Pattern Recognition, vol. 129, art. 108766, 2022, doi: 10.1016/j.patcog.2022.108766.

[8] U. Ali et al., "Leveraging Deep Learning with Multi-Head Attention for Accurate Extraction of Medicine from Handwritten Prescriptions," in 2024 International Conference on Robotics and Automation in Industry (ICRAI), 2024, doi: 10.1109/ICRAI62391.2024.10894218.

[9] S. J. Reddy et al., "Multimodal Deep Learning with XAI for Transparent and Accurate Doctor Prescription Recognition," in 2024 International Conference on Emerging Research in Computational Science (ICERCS), 2024, doi: 10.1109/ICERCS63125.2024.10895660.

[10] G. Pavithiran et al., "Doctor's Handwritten Prescription Recognition System in Multi-Language Using Deep Learning," arXiv:2210.11666, 2022 (preprint).

[11] D. Roja Ramani et al., "PrescriptIQ: Revolutionizing Healthcare with AI-Powered Multilingual Prescription Decoding," in 2024 5th International Conference on Image Processing and Capsule Networks (ICIPCN), 2024, doi: 10.1109/ICIPCN63822.2024.00070.

[12] N. M. Jyothi, J. B. Simha and K. V. Kiran Kumar, "Word Prediction from Medical Prescription via Transfer Learning with Pre-trained CNN, GAN and BiLSTM Integration," in 2024 2nd International Conference on Networks, Multimedia and Information Technology (NMITCON), 2024, pp. 1–7, doi: 10.1109/NMITCON62075.2024.10698946.

[13] S. Chowdhury and K. K. Mandal, "A Comprehensive Review on Handwritten Medical Prescription Recognition Using Deep Learning Techniques," in Emerging Trends in Smart Computing, Communication and Automation, CRC Press, 2026.

[14] K. K. Mandal and A. C. Mondal, "Proposed Privacy-Preserving Framework on Medical Handwritten Prescription Using Blockchain," in AI in Smart and Secure Healthcare, Springer Optimization and Its Applications, vol. 243, pp. 361–377, 2026.