Int. J. Sci. R. Tech., 2026 3(11)
A Multidisciplinary peer-reviewed Journal
www.ijsrtjournal.com [ISSN: 2394-7063]

Original Article

Deep Learning-Based Optical Character Recognition for Automated Medical Prescription Decoding: A Critical Review of Architectures, Techniques, and Clinical Applications

Simranjeet Kaur, Shalu Gupta*

Dept. of Computer Applications, Guru Kashi University, Punjab, India.

ABSTRACT

The handwritten medical prescription decoding is a multi-stage problem of document-intelligence of which requires visual recognition, layout interpretation, medical terminology and clinically responsible verification tasks. This critical review summarizes deep-learning architectures and integrated document-processing approaches that are relevant to the recognition of prescriptions, with a particular focus on the differences between image-to-text transcription and the extraction of meaningful information for medications from the documents. An analysis of convolutional neural networks (CNNs), recurrent and bidirectional sequence models, convolutional recurrent neural networks (CRNNs), connectionist temporal classification (CTC), attention mechanisms, Transformer based OCR, layout-aware document models and vision-language approaches. It is found that the results in medicine-name transcription, region detection, multilingual processing, and structured record generation are progressing with prescription-specific studies, but that there are differences between the sets of medicine names, languages, recognition units and evaluation protocols used in the studies. Based on the review, the authors suggest a framework with human in the loop that connects image quality control, region detection, visual encoding, sequence recognition, clinical entity extraction, terminology validation, uncertainty estimation, professional verification and secure integration into health records. Analysis reveals that public data sets suffer from various shortcomings, such as lack of writer independence, language coverage, field-level error reporting and prospective clinical assessment. It concludes that a high OCR is not sufficient to ensure reliable prescription decoding - it's essential that there is visual evidence that can be traced, uncertainty that can be calibrated, clinically meaningful evaluation and qualified human oversight.

Keywords: Handwritten medical prescriptions, handwritten text recognition, optical character recognition, image preprocessing, deep learning, CNN, medical document analysis, medication safety

INTRODUCTION:

Prescriptions are still written on paper in clinical settings where electronic prescriptions are not fully implemented or are still in addition to handwritten ones. A prescription might include the name of the drug, dosage, form, route, frequency/duration, quantity, and instructions. The elements are typically presented in an irregular format with compact characters or sometimes abbreviations, symbols and numbers, as well as printed and handwritten text. Automated decoding is then no longer similar to a generic OCR, needs to be accurate, preserve the document structure and correctly identify the name of the medicines with its instructions.

The main technical challenge is the visual ambiguity in combination with domain specific language. Some characters may be hard to tell apart due to connected strokes, spacing inconsistencies, slant, poor mobile images, and overlapping strokes. The names of the drugs may not be familiar or may look alike, and the decimal points and units along with the expressions for dosage may be critical. While it is important to have visual evidence, a language model can also generate a plausible drug name if the evidence in the image is light.

Deep learning has changed text recognition from the hand-crafted features and explicit segmentation to learned visual representations and sequence modelling. CNNs have been used to extract local stroke

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

patterns; recurrent networks are used to model ordered dependencies; CRNN-CTC systems are used to learn image-to-sequence mappings without character-level segmentation; and Transformers are attention-based alternatives that perform well in the context of modelling.CNNs have been used to extract local stroke patterns; recurrent networks are used to model ordered dependencies; CRNN-CTC systems are used to learn image-to-sequence mappings without character-level segmentation; and Transformers are attention-based alternatives with good modelling of the context [1]–[8]. Document models include LayoutLM [1] and LayoutLMv3 [2] which use spatial and visual context, and OCR-free models including Donut [9]–[11] which look at direct image-to-structured-output generation.

The aim of this article is not to review in general the methods and pre-processing used for the whole area of speech processing. Its contribution is a synthesis, which is architecture-centred and pipeline-oriented: it compares model families for the recognition of different types of prescriptions, looks at evidence for a prescription and compares that with the clinical entity to be recognized; it distinguishes OCR from clinical entity understanding; and proposes an integrated human-supervised framework for the decoding of prescription.

Review Questions and Literature-Search Strategy

(RQ1) What are the differences between CNN, CRNN, handwritten prediction recognition network (BiLSTM/CTC), attention and Transformer architecture in handwritten prescription recognition? (RQ2) What evidence have been reported from the various studies of specific prescriptions and what impedes making comparisons across these studies? (RQ3) What would be the best way to take a recognized text and make it structured medication data and not be overpowered by semantic corrections over the image evidence? (RQ4) What assessment and stewardship considerations need to be taken into account prior to implementation in clinical or pharmacy processes?

Search sources and search terms

The literature search protocol will include studies that are specific to the drug and also methods. The planned bibliographical sources include Scopus, Web of Science, IEEE Xplore, PubMed/MEDLINE, ACM Digital Library and Google Scholar. Bibliographic metadata is verified by publisher records and landing pages of the DOIs. Some examples of search strings are: ("handwritten medical prescription" OR "doctor prescription" OR "medicine name recognition") AND (OCR OR "handwritten text recognition" OR CRNN OR BiLSTM OR CTC OR TrOCR OR Transformer); (prescription AND ("document understanding" OR "medication extraction" OR "clinical NLP")); and ("medical document" AND (layout OR segmentation OR "vision-language model")). Some of the basic architectures that are used in searching for the architecture of foundations include: "connectionist temporal classification", "convolutional recurrent neural network", "Transformer OCR", and "document image understanding".

The current manuscript is not a completed PRISMA systematic review or meta-analysis, but is a critical narrative review of the literature that is still structured. The evidence window highlights the timeframe of November 2017 – 18, but earlier foundational studies that explain key architectures or datasets are retained if they explain essential architectures or datasets. The manuscript does not report the numbers retrieved from the database, the number of duplicates and the numbers of PRISMA flow as a complete databases export and screening log was not available in the source draft. These must be added in after the authors have performed and documented searches in each of the named databases and not be made up totals.

Eligibility criteria and data extraction

Inclusion criteria	Exclusion criteria
Peer-reviewed journal or conference studies on handwritten/ mixed prescription recognition, medicine extraction, prescription layout analysis, or directly relevant HTR/document-AI methods.	Studies limited to unrelated medical imaging, generic printed OCR without transfer relevance, or administrative documents unrelated to prescriptions.
Foundational architecture papers and datasets needed to explain CNN, RNN/LSTM, CTC, Transformer, and document-understanding methods.	Promotional product descriptions, duplicate records, and sources lacking enough technical information to interpret the method.

Inclusion criteria	Exclusion criteria
Studies reporting a dataset, model design, evaluation measure, or clinically relevant system outcome.	Claims based only on unsupported secondary summaries or unverifiable performance statements.

The review lists the tasks, data source, language/script being used, recognition unit (word, line, region, or entire page), the model family, the preprocessing, decoding/post-processing, the evaluation metrics, the validation design and the limitations for each empirical study. The studies are judged qualitatively according to the following criteria: dataset transparency, writer/document separation, reproducibility, metric appropriateness, external validation and clinical relevance. The literature does not necessarily conform to a single format to be included and, therefore, the recognition scores are not combined and the studies are not ranked by the accuracy of the headlines.

Architecture Foundations and Critical Comparison

CNN-based visual encoding

A CNN is used to decompose the image pixels into hierarchical feature maps. Early convolutional layers often represent local edges, stroke direction and texture; deeper layers are able to integrate the two into more discriminative visual patterns. In sequence-recognition systems, the horizontal dimension of the feature map could be a sequence of feature vectors. CNNs are thus good visual encoders, but do not necessarily represent the dependencies between characters in a word or line. CNN features may be sensitive to the resolution, blur, stroke thickness and preprocessing decisions for prescription images. They are more often used as the visual interface of a sequence recogniser than as a full-blown prescription-decoder.

RNN and BiLSTM sequence modelling

Recurrent neural networks take feature vectors, and process them one by one. LSTM units maintain memory over longer periods of time and bidirectional LSTMs take the sequence both forward and backward. This gives both sides of the character for each output position, so that it can help to disambiguate ambiguous handwritten characters. Graves et al. proposed the recurrent neural network for handwriting recognition without any constrains [3]. However, recurrent models deal with the positions of the sequences serially, this makes them less parallel, and context can provide evidence of a wrong visual interpretation, if handwriting is very ambiguous.

CRNN and CTC

A CRNN is the combination of convolutional feature extraction and recurrent sequence modelling and a transcription layer. The CTC training objective is to map a variable length image sequence to text, instead of having to manually align each character to a specific position in the image [1] and [2]. The CTC is the sum of the probabilities of the paths that survive the blank removal and repeated label merging processes at the frame level that lead to the target sequence. This is good for cursive handwriting as character boundaries are not clearly defined. The monotonic correspondence between the input sequence and the output labels is however assumed and it is difficult to train CTC if the text reading order is not a straightforward sequence from left to right.However, CTC requires a monotonic correspondence between the input sequence and the output labels, and it is difficult to train CTC if the text reading order is not a simple left-to-right sequence.

Attention and Transformer-based OCR

The recognition process can attend to various visual regions during the generation of text in attention-based encoder–decoder models. Transformers dispense with recurrent sequence processing in favor of self-attention which allows parallel representation learning and modelling of long-range relationships [7]. TrOCR consists of an image Transformer encoder and a text Transformer decoder pre-trained, and can

be fine-tuned for handwritten recognition [8]. In situations where the prescription-specific labelled data is less, this pretraining strategy is important. But, Transformers can be quite resource intensive and need to be carefully adapted to the domain. Their ability to generate language models can also be exploited to produce fluent tokens without accompanying source visuals, making it imperative to have confidence estimation and source-image review.

Layout-aware and OCR-free document models

The key difference between LayoutLM and LayoutLMv3 is the introduction of unified text-image pretraining objectives for document AI in LayoutLMv3 [9], [10] that jointly represent text and spatial layout. These work where the meaning of the medicine name, strength or instruction is dependent on its position. Donut aims to directly produce structured output from document images, without the need for OCR, by tackling the problem of OCR-free document understanding [11]. Depending on the models, these models can eliminate the need for OCR and layout steps, but there is a need for prescription-specific validation. The model trained on forms, receipts or printed documents cannot be assumed to generalize to locally variable layouts or multilingual prescriptions or to cursive handwriting.

Architecture	Primary role	Strengths for prescriptions	Limitations / risks	Best-fit use
CNN	Visual feature extraction	Learns stroke, edge, and local shape features; efficient and widely reusable.	Does not independently model long character sequences or document semantics; sensitive to image quality.	Feature encoder within CRNN or hybrid models.
BiLSTM / RNN	Sequential context modelling	Uses left and right sequence context; supports variable-length handwriting.	Sequential computation; limited global document context; can propagate mistaken context.	Word- or line-level sequence modelling.
CRNN–CTC	End-to-end image-to-text sequence recognition	Avoids explicit character segmentation; established HTR design; practical for line images.	Monotonic alignment assumptions; limited layout understanding; decoding errors remain possible.	Handwritten word/line transcription.
Attention / Transformer OCR	Global visual and linguistic context	Captures long-range dependencies; pretrained encoders/decoders can reduce task-specific data needs.	Higher compute/data demands; hallucinated or language-prior-driven substitutions require safeguards.	Fine-tuned prescription word/line recognition.
Layout-aware / OCR-free document Transformer	Document-level structure and field extraction	Uses visual and spatial relationships; can produce structured fields.	Requires representative layouts and strong annotations; handwriting performance may vary.	Full-page parsing and structured prescription extraction.

Prescription-Specific Evidence

Studies of the effects of prescription HTs are less common and more varied than general HT studies. Rani et al. studied the online-stylus-based medicine name recognition from handwriting/signatures with a small sized medicine-name dataset and achieved the recognition accuracy that varies according to the user [18]. This setting is different from paper based face-to-face recognition of paper prescriptions; the results should not be used to directly compare image-based OCR.

In this work, Zaman et al. took 893 annotated Bangladeshi prescription images as a test set and tried to use TrOCR to solve the problem. Prior to their being corrected by edit distance, they reported 16% CER and 36% WER, after editing 7% CER and 11% WER [19]. The study demonstrates the capability of pretrained Transformer OCR, as well as the efficacy of domain-specifically correcting the recognition results, yet it also indicates the need for reporting raw and corrected recognition results separately.

For medicine extraction, Ali et al. used Mask R-CNN region segmentation and TrOCR, multi-head attention and positional embeddings to achieve a CER of 1.4% on their evaluation [20]. Datta et al. proposed a wider prescription information pipeline that has ROI detection accuracy of 99.6% and spelling correcting accuracy of 96% using YOLO-based ROI detection, OCR, spelling correction and structured medication information [21]. Ullah et al. presented an approach using TrOCR and Roboflow and achieved 8.2% CER,

12.5% WER and 81.3% exact-match accuracy [22]. Note: Do not use these values as a leader board as they stem from different data sources, tasks, recognition units and evaluation procedures.

There has been some recent transcription work done which is beyond that. Biswas et al. propose a deep-neural solution to doctor's handwritten prescription recognition and generation of digital patient records [23]. Ameer et al. suggest a hybrid pipeline which categorizes the image of the prescription and directs it for OCR engines and compares it with a vision-language model [24]. Polisetty et al. tackle the problem of treatment-regimen segmentation with an attention-gated U-Net architecture with their report of IoU and Dice measures for segmentation, instead of the text recognition task [25]. Bukhari et al. propose a real-world handwritten clinical prescription dataset for detection and recognition, BHCP-DOC, that is bilingual [26]. These studies all corroborate a pipeline approach where detection of a region, recognition of the text contained in this region, segmentation and structuring of the clinical text are separate but related tasks.

Study	Task / approach	Reported evidence	Interpretation and limitation
Rani et al. (2022) [18]	Online stylus-based medicine-name recognition; signature-verification features.	Reported user-dependent accuracies; highest reported value 84%.	Online handwriting differs from offline photographed prescriptions; small dataset.
Zaman et al. (2024) [19]	TrOCR on 893 Bangladeshi prescription images; edit-distance correction.	CER/WER 16%/36%; after correction 7%/11%.	Useful Transformer case study; correction must be assessed separately from raw OCR.
Ali et al. (2024) [20]	Mask R-CNN + TrOCR + multi-head attention for medicine extraction.	Reported CER 1.4%.	Dataset and protocol are study-specific; not directly comparable with other CER values.
Datta et al. (2025) [21]	YOLO ROI detection, OCR, spelling correction, structured EHR data.	Reported ROI detection accuracy 99.6%; spelling-correction accuracy 96%.	Component-level measures do not equal end-to-end clinical correctness.
Ullah et al. (2025) [22]	TrOCR-based handwritten prescription OCR.	Reported CER 8.7%, WER 12.5%, exact match 81.3%.	Online-first study; external validation and dataset transfer remain important.
Biswas et al. (2026) [23]	Deep neural prescription recognition and digital record generation.	Prescription-specific system reported.	Do not infer generalization beyond the study's dataset and protocol.
Ameer et al. (2026) [24]	Hybrid image classification and OCR routing, benchmarked against GPT-4o.	Reports an adaptive prescription digitization pipeline.	Mixed prescription-image scope and default OCR engines constrain direct interpretation.
Polisetty et al. (2026) [25]	Attention-gated U-Net for treatment-regimen segmentation.	Reported IoU 87.2% and Dice 92.9%.	Segmentation metrics are not OCR accuracy.
Bukhari et al. (2026) [26]	BHCP-DOC bilingual handwritten clinical prescription dataset.	Dataset resource for detection and recognition.	Dataset availability supports benchmarking; independent evaluation is still required.

Integrated Prescription-Recognition Framework

The stages in the prescription-decoding systems are often confused, as shown in Figure 1. First, the image is captured and assessed for blurriness, skewing, resolution, lighting and any information that is of a sensitive nature to privacy. Secondly, preprocessing and layout analysis are used to identify important text areas, without altering the original image. Third, a CNN feature extractor extracts features from the visual data. Fourth, a sequence recognizer that can be a CRNN–BiLSTM–CTC or a fine-tuned Transformer produces one or more text hypotheses. Fifth, clinical NLP gets the names of the medicines, strength, dosage form, route, frequency, duration and instructions. Sixth, the use of terminology resources could aid in normalisation, but any correction should be explainable in terms of the visual evidence [16]. Finally, fields with uncertainty or risk of export to an incorrect patient are forwarded to either a pharmacist or clinician for confirmation prior to the structured export [17].

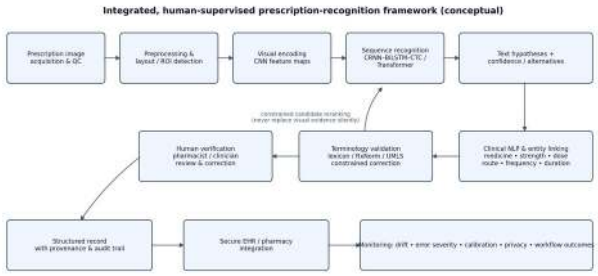

Figure 1: ***Conceptual integrated framework for human-supervised handwritten prescription recognition and structured information extraction.***

The framework is only conceptual and not as an experimentally tested system. The main design principle is to not override the information from the visual evidence with a semantic plausibility. The OCR output (raw OCR), corrected candidate, confidence score, source image region and human correction should be kept as separate fields.

Clinical NLP, Terminology and Semantic Validation

Currently, the output of OCR is not structured prescription. In clinical NLP, it is necessary to recognize entities like medicine name, strength, dosage form, route, frequency, duration, quantity and instruction, and to link them with the correct medicine. There are also biomedical language models (BioBERT) [12] and biomedical domain-specific models (PubMedBERT) [13] that represent biomedical text tasks. They can be helpful in the tasks of named entity recognition and relation extraction but may be confused by noisy OCR strings and local prescription abbreviations, and need to be task validated.

Concept mapping can be assisted by terminology resources like the Unified Medical Language System, and the use of synonyms can also be assisted [14]. Normalization in practice should be limited: If the visual evidence, the context, or attributes of the relevant terminology are not sufficiently coherent, the candidate medicine name should not be linked to the terminology entry. Lexicon match is not necessarily a measure of correct reading of a token. This is especially true of names of medicines that look similar, the strength of their medicine in decimal form, the units and abbreviations used.

The raw visual transcription is preserved, as is an alternative recognition hypothesis, any normalized form, the evidence and the rule by which it was normalized, and the final correction by the transcriber. This approach will make auditing easier and lower the possibility of a language model making a "best guess" from a picture and creating a false medication log.

Evaluation: From OCR Accuracy to Clinical Correctness

Character error rate (CER) and word error rate (WER) are used for measuring the differences in transcription on the basis of edit distance. They have been helpful for comparing text recognition in the same protocol, but neither of them has any clinical meaning to signify the severity of an error. The Complete medicine-name or dosage is informative for Exact Match accuracy and the extracted entities and relations are suitable for precision, recall and F1. The measures to assess the quality of region detection and segmentation include intersection over union (IoU) and Dice score. In addition to deployment studies, latency, memory usage, dependency robustness to writers and image conditions, confidence calibration, abstention and workload for human reviewers should be reported.

Evaluation layer	Recommended measures	Why it matters
Text transcription	CER, WER, exact-match accuracy	Measures character-, word-, and complete-string recognition.
Medication entities	Precision, recall, F1; exact match by field	Distinguishes medicine names from strengths, units, and instructions.
Layout / region detection	IoU, Dice, precision/recall	Measures whether the correct prescription regions are located.
Clinical safety	Error severity; high-risk field error rate; correction rate	Separates harmless spelling differences from medication-relevant errors.
Reliability	Confidence calibration, abstention rate, selective risk	Tests whether uncertain predictions are escalated appropriately.
Efficiency	Latency, throughput, memory, model size, hardware	Supports realistic deployment decisions.
Generalization	Writer-independent split; external institution; subgroup metrics	Tests transfer to unseen writers, hospitals, scripts, and devices.

Make conservative claims when cross-study. This CER of 1.4% on a medicine-name crop cannot be directly compared to a WER of 12.5% for full prescription lines, or to a 96% accuracy of spelling corrections. To have any numerical ranking meaning, the unit of analysis, the composition of the data, the language, the separation of the writers, the annotation policy or the evaluation code should be comparable.

Datasets and Generalization

While general handwriting databases, e.g., IAM, can be employed for pre-training and method development, they are not of the same vocabulary, abbreviations, layouts or dosage expressions that are used in prescriptions. Prescription datasets are thus of vital importance. RxHandBD supplies a set of word-image data for handwritten medicine words [15] and BHCP-DOC supports real-world clinical prescription texts in the target languages, English and Chinese [26]. Prescription datasets should include information about source institution, language/script, number of writers, number of documents, conditions of image acquisition, recognition unit, policy for annotations, licensing, train/validation/test partitioning.

If identifiers are available, it is better to use writer-independent and/or document-independent evaluation. Optimistic estimation can be obtained by randomly splitting images, which can result in the same prescription author or words being separated between training and test partitions. To assess whether a model can be transferred to new environments, such as across hospitals, regions, and specialties, and across acquisition devices, it is necessary to undergo external evaluation.

Challenges and Future Research

The main outstanding research challenges are: (1) lack of representative and public prescription datasets; (2) lack of coverage on handwriting in multiple languages and code-mixed; (3) lack of external and writer-independent validation; (4) the fact that the CER/WER is used without taking into consideration the severity of the errors of the field; (5) absence of uncertainty calibration and abstentions; (6) absence of full-page layout and relation modelling; (7) privacy and interoperability constraints; and (8) lack of prospective workflow evaluations.

Future studies could build multi-institutional datasets that are de-identified and consistently annotated and licensed, investigate the performance of CNN–BiLSTM–CTC and Transformer recognizers on common splits, and split the raw OCR from the lexicon-corrected output; evaluate medicine, strength, and dosage fields separately; and report high-risk error rates. A range of methods to mitigate data scarcity, including self supervised learning, synthetic handwriting, few-shot adaptation and federated learning, may be useful, but these should be shown to be effective on an external test set. Explicit uncertainty handling should be used to constrain vision-language models and strong OCR baselines should be used to evaluate them. The extent of review time, the amount of corrections that must be made, how much patients rely on the OCR and the safety outcomes should be measured in clinical studies, not assuming that the more patient safe the better the OCR will be.

CONCLUSION

Prescription decoding using deep-learning techniques has evolved from text recognition to comprehensive document recognition and information extraction of medication. CNNs add visual features, BiLSTM and CRNN-CTC models are able to transcribe sequences, Transformers (e.g., TrOCR) are able to recognize sequences in a contextual way; and layout-aware or vision-language models take the task to the document level and structure. Numerical results for promising approaches in prescription specific studies are hard to compare as there are different datasets, languages, tasks and metrics. A clinically responsible system must maintain the evidence of the image, differentiate between transcription and semantic normalisation, show uncertainty for ambiguous or high-risk fields and call in qualified human verification of ambiguous or high-risk fields. This proposed framework is meant to be a conceptual synthesis that will provide a direction for research and evaluation, but not an experimental validation.

REFERENCES

1. A. Graves, S. Fernández, F. J. Gomez, and J. Schmidhuber, “Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks,” in Proc. 23rd Int. Conf. Machine Learning, 2006, pp. 369–376, doi: 10.1145/1143844.1143891.

2. B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 11, pp. 2298–2304, 2017, doi: 10.1109/TPAMI.2016.2646371.
3. A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber, "A novel connectionist system for unconstrained handwriting recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 5, pp. 855–868, 2009, doi: 10.1109/TPAMI.2008.137.
4. U.-V. Marti and H. Bunke, "The IAM-database: An English sentence database for offline handwriting recognition," Int. J. Document Anal. Recognit., vol. 5, pp. 39–46, 2002, doi: 10.1007/s100320200071.
5. W. AlKendi, F. Gechter, L. Heyberger, and C. Guyeux, "Advancements and challenges in handwritten text recognition: A comprehensive survey," J. Imaging, vol. 10, no. 1, Art. no. 18, 2024, doi: 10.3390/jimaging10010018.
6. C. Garrido-Munoz, A. Rios-Vila, and J. Calvo-Zaragoza, "Handwritten text recognition: A survey," IEEE Trans. Pattern Anal. Mach. Intell., 2026, doi: 10.1109/TPAMI.2025.3646002.
7. A. Vaswani et al., "Attention is all you need," in Adv. Neural Inf. Process. Syst., vol. 30, 2017.
8. M. Li et al., "TrOCR: Transformer-based optical character recognition with pre-trained models," in Proc. AAAI Conf. Artificial Intelligence, vol. 37, no. 11, pp. 13094–13102, 2023, doi: 10.1609/aaai.v37i11.26538.
9. Y. Xu, M. Li, L. Cui, S. Huang, F. Wei, and M. Zhou, "LayoutLM: Pre-training of text and layout for document image understanding," in Proc. ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining, 2020, pp. 1192–1200, doi: 10.1145/3394486.3403172.
10. Y. Huang, T. Lv, L. Cui, Y. Lu, and F. Wei, "LayoutLMv3: Pre-training for document AI with unified text and image masking," in Proc. ACM Int. Conf. Multimedia, 2022, pp. 4083–4091, doi: 10.1145/3503161.3548112.
11. G. Kim et al., "OCR-free document understanding transformer," in Computer Vision – ECCV 2022, 2022, pp. 498–517, doi: 10.1007/978-3-031-19815-1_29.
12. J. Lee et al., "BioBERT: A pre-trained biomedical language representation model for biomedical text mining," Bioinformatics, vol. 36, no. 4, pp. 1234–1240, 2020, doi: 10.1093/bioinformatics/btz682.
13. Y. Gu et al., "Domain-specific language model pretraining for biomedical natural language processing," ACM Trans. Comput. Healthcare, vol. 3, no. 1, Art. no. 2, 2021, doi: 10.1145/3458754.
14. O. Bodenreider, "The Unified Medical Language System (UMLS): Integrating biomedical terminology," Nucleic Acids Res., vol. 32, suppl. 1, pp. D267–D270, 2004, doi: 10.1093/nar/gkh061.
15. S. Sharjeel and M. U. Arif, "AI-based OCR system for handwritten medical prescription recognition and interpretation," Indian J. Comput. Sci. Technol., vol. 4, no. 2, pp. 365–370, 2025, doi: 10.59256/indjcst.20250402050.
16. T. Mankash, V. S. C. Kota, A. De, P. Prakash, and K. Jadhav, "MIRAGE: Multimodal identification and recognition of annotations in Indian general prescriptions," arXiv:2410.09729, 2024.
17. D. W. Bates and H. Singh, "Two decades since To Err Is Human: An assessment of progress and emerging priorities in patient safety," Health Affairs, vol. 37, no. 11, pp. 1736–1743, 2018, doi: 10.1377/hlthaff.2018.0738.
18. S. Rani, A. U. Rehman, B. Yousaf, H. T. Rauf, E. Abouel Nasr, and S. Kadry, "Recognition of handwritten medical prescription using signature verification techniques," Comput. Math. Methods Med., vol. 2022, Art. no. 9297548, 2022, doi: 10.1155/2022/9297548.
19. M. F. U. Zaman, B. Rahman, A. Rubyat, N. Halder, A. Mahmud, A. Islam, and M. A. Amin, "Recognizing medicine names from Bangladeshi handwritten prescription images using TrOCR," in Proc. 7th Asia Conf. Cognitive Engineering and Intelligent Interaction, 2024, pp. 1–5, doi: 10.1109/CEII65291.2024.00020.
20. U. Ali, S. Ranmbail, M. Nadeem, H. Ishfaq, M. U. Ramzan, and W. Ali, "Leveraging deep learning with multi-head attention for accurate extraction of medicine from handwritten prescriptions," in Proc. Int. Conf. Robotics and Automation in Industry, 2024, doi: 10.1109/ICRAI62391.2024.10894218.

21. A. Datta, M. M. Hasan, N. Mahmud, B. J. Ferdosi, R. Islam, Z. Rahman, and S. T. Mehedi, "Preserving medical information from doctor's prescription ensuring relation among the terminology," Comput. Biol. Med., vol. 188, Art. no. 109812, 2025, doi: 10.1016/j.compbiomed.2025.109812.
22. H. Ullah, M. Tanveer, and A. Jan, "Enhancing handwritten prescription recognition with AI-driven OCR," J. Comput. Biomed. Inform., vol. 9, no. 2, 2025, doi: 10.56979/902/2025.
23. B. Biswas, B. K. Sarkar, A. Mustafi, and A. Mitra, "A novel approach of doctors' handwritten medical prescription recognition using a deep neural network and generating digital records of the patients," J. Dyn. Games, 2026, doi: 10.3934/jdg.2026016.
24. O. Z. Ameer, M. H. Majeed, S. L. Kailan, S. M. B. Alsabti, and A. M. Elhenidy, "A hybrid deep learning-based OCR model for handwritten medical prescriptions," Front. Med., vol. 13, Art. no. 1856485, 2026, doi: 10.3389/fmed.2026.1856485.
25. R. G. R., S. S., and V. N. M. Aradhya, "Treatment regimen segmentation from handwritten medical prescriptions using advanced neural network," Artif. Intell. Appl., 2026, doi: 10.47852/bonviewAIA62027648.
26. F. K. Bukhari, M. Coustaty, M. M. Luqman, S. Bakkali, M. G. Madden, and N. I. Kajla, "BHCP-DOC: A real-world bilingual dataset for handwritten clinical prescription detection and recognition," in Pattern Recognition. ICPR 2026 International Workshops, 2026, pp. 448–463, doi: 10.1007/978-3-032-39379-1_31.

HOW TO CITE: Simranjeet Kaur, Shalu Gupta, Deep Learning-Based Optical Character Recognition for Automated Medical Prescription Decoding: A Critical Review of Architectures, Techniques, and Clinical Applications, Int. J. Sci. R. Tech., 2026, 3 (11), 1395-1403. https://doi.org/10.5281/zenodo.24556339