Received: 12 September 2025 | Revised: 30 December 2025 | Accepted: 14 January 2026 | Published online: 25 January 2026

RESEARCH ARTICLE

Artificial Intelligence and Applications
2026, Vol. 00(00) 1–12
DOI: 10.47852/bonviewAIA62027648

BON VIEW PUBLISHING

Treatment Regimen Segmentation from Handwritten Medical Prescriptions Using Advanced Neural Network

Rekha G. R.[1], **Siddesha S.**[1,*], **and V. N. Manjunath Aradhya**[1]

[1] *Department of Computer Applications, JSS Science and Technology University, India*

Abstract: Handwritten medical prescriptions are a critical yet under-digitized component of clinical workflows, often serving as a source of ambiguity due to illegible handwriting, overlapping text blocks, and structural inconsistencies. The automatic segmentation of such prescriptions into meaningful textual blocks is vital for downstream tasks like drug recognition and dosage extraction. Traditional methods grounded on connected components or projection profiles often falter under the irregularities of freeform handwriting. To address these limitations, the paper proposes an advanced deep learning architecture—PrescNet—that primarily segment the treatment regimen (medicine and its associated components) as text-blocks using classical U-Net design with spatial–channel attention gates and a lightweight 32 channel projection layer to better capture salient features in prescription images. The model is trained on a custom dataset with pixel-level annotations and evaluated using 10-fold cross-validation with varying data splits. Experimental results demonstrated that the proposed architecture significantly transcend the baseline variants and a few state-of-the-art deep learning models of text-line segmentation achieving an Intersection over Union (IoU) of 87.2%, Dice score of 92.9%, and a minimal Dice loss of 0.071. The results validate its effectiveness in handling complex handwritten layouts, establishing its suitability for real-world clinical applications.

Keywords: handwritten medical prescription, semantic block segmentation, deep learning, U-Net, attention mechanism

1. Introduction

The digitization of handwritten medical prescriptions remains a complex yet crucial component of modern healthcare systems. Accurate extraction and structural preservation of prescription content is vital, particularly in contexts where the spatial relationship between drug names and associated information (e.g., dosage, frequency, and administration route) must be maintained to avoid misinterpretation. While recent years have witnessed a surge in digitization frameworks targeting hospital records and patient documentation [1, 2], the challenge of segmenting unstructured, handwritten prescriptions at the text-block level has not been adequately addressed.

Many existing literatures focus on either document-level digitization [3] or extraction of structured data from typed or form-based records [4, 5]. However, handwritten prescriptions are inherently irregular, featuring highly varied writing styles, non-linear layouts, overlapping components, and skewed text lines. These characteristics pose a significant challenge to conventional optical character recognition (OCR) and document parsing techniques, which often rely on rigid assumptions regarding textual structure [6].

Emerging solutions leveraging deep learning have demonstrated improved capability in handling such complexities. Models combining convolutional neural networks with sequence-to-sequence learning or hybrid approaches incorporating attention mechanisms have shown promise in historical manuscript segmentation [7]. Yet, these methods primarily target line-wise or entity-level segmentation, insufficient for preserving the semantic linkage within prescription blocks. Segmenting text blocks, as opposed to lines, become essential when the goal is to preserve the relationship between a medicine name and its modifiers—a relationship that may span multiple spatial zones within a handwritten layout.

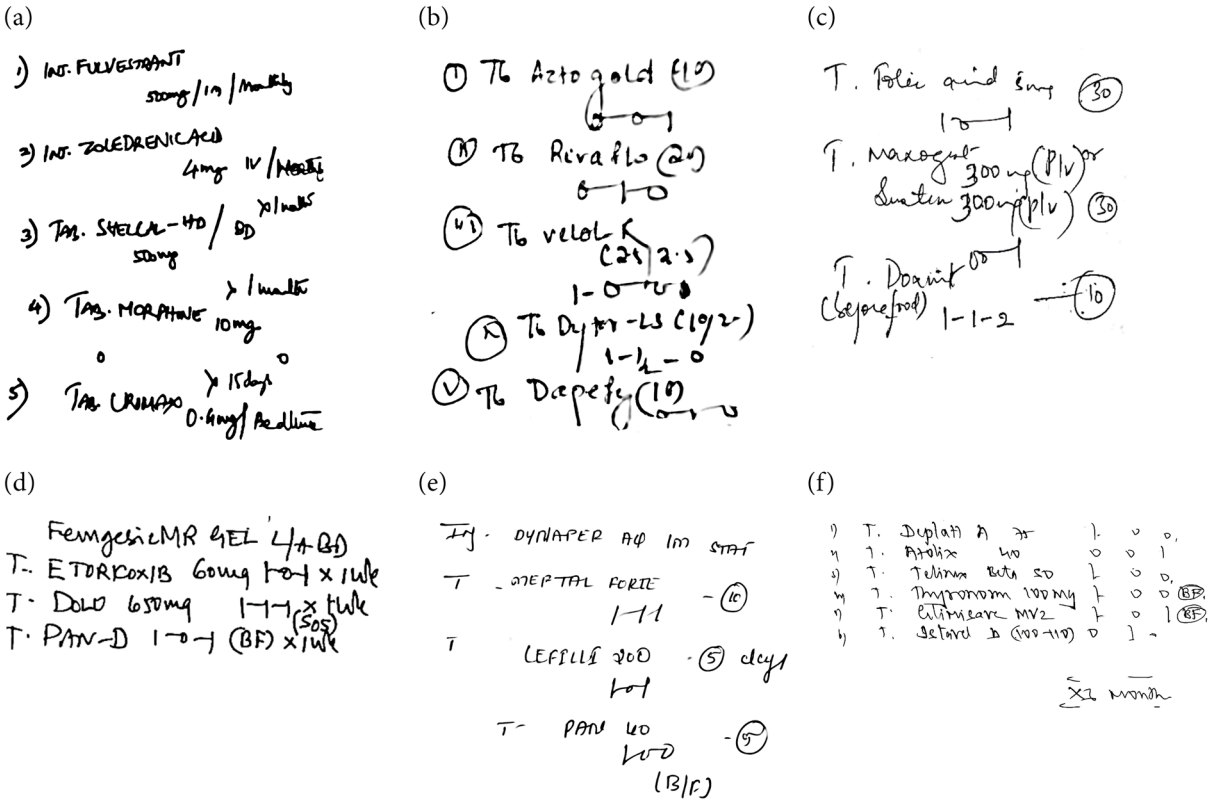
In recent work, machine learning-driven digitization pipelines have begun to bridge the gap between image understanding and clinical documentation by adopting structure-aware recognition techniques [8]. These methods underscore the necessity of not only recognizing text but also retaining its contextual and structural relevance—particularly in handwritten clinical content where visual cues often dictate semantic interpretation.

Figure 1 shows a few samples of region of interest (ROI) extracted handwritten medical prescriptions. Each prescription demonstrates a different mode of prescribing the medicines. The medicine and its associated components in few prescriptions are distributed across two or three lines (as given in Figure 1(a), (b), (c), and (e)). In certain instances, like Figure 1(e), the advice is given in a single line as well as in multi-lines. Few cases show orientations and overlapping characters (Figure 1(a), (d), and (f)). Hence, to preserve the semantic meaning of the medicines along with its components, text-block segmentation is attained in the current work instead of regular text-line segmentation. Consequently, this work proposes a novel deep learning-based segmentation model tailored for the specific task of handwritten prescription block extraction. This block-level segmentation built upon the foundational U-Net architecture and amplifies with spatial and channel-wise attention mechanisms. Thus, the proposed PrescNet aims to enhance the accuracy of block segmentation while preserving the spatial context essential for downstream information extraction. The model is empirically validated on a custom dataset of prescriptions, as no standard datasets are publicly available and its performance is benchmarked against multiple deep learning models as well as U-Net variants across different data splits and k-fold cross-validation settings.

By addressing a previously underexplored granularity of prescription segmentation, this research contributes a robust framework

***Corresponding author:** Siddesha S., Department of Computer Applications, JSS Science and Technology University, India. Email: siddesh.shiv@jssstuniv.in

Figure 1
Different samples of handwritten medical prescriptions



for structured prescription digitization—paving the way for downstream clinical NLP applications, safer electronic medical record integration, and enhanced patient care automation.

1.1. Key contributions

The core contributions of this study include:

1) Reformulation of handwritten prescription that primarily highlights the block-level segmentation task to preserve semantic and structural coherence of medical components instead of regular text-line segmentation.
2) Proposed PrescNet, an advanced encoder–decoder architecture that integrates spatial and channel attention with attention-gated skip connections and a 32-channel projection layer for enhanced prescription segmentation.
3) Extensive benchmarking against three state-of-the-art models and four U-Net variants and evaluate performance across different data splits and 10-fold cross validation to ensure robustness and generalizability.
4) Detailed error analysis on overlapping and ambiguous text regions.
5) Demonstration of the model's real-world applicability for digitizing handwritten prescriptions in clinical workflows.

2. Literature Review

2.1. Traditional text-line and block segmentations

Text-line and block segmentations have long served as foundational steps in document image analysis, particularly in the processing of historical and handwritten manuscripts. Early approaches relied heavily on projection profiles, connected component analysis, and morphological operations. For instance, Wahl et al. [9] proposed one of the earliest frameworks for block segmentation and text extraction in mixed text-image documents using projection-based heuristics. However, the method often struggled with irregular line spacing, overlapping characters, and skewed writing.

To address these limitations, Li et al. [10] introduced a script-independent segmentation technique designed for freestyle handwritten documents, which leveraged directional continuity for line detection. While these methods improved flexibility, they remained limited in their ability to handle dense layouts and multi-touching text instances. More recent efforts have introduced neural and hybrid techniques. For example, Vadlamudi et al. [11] incorporated recognition-based evaluation and high-precision attention mechanisms respectively, but still primarily addressed line-wise separation.

Traditional segmentation methods also struggled in complex clinical contexts where handwriting irregularities are frequent. Shivakumara et al. [12] explored segmentation in struck-out text, a scenario common in clinical revisions, but their technique is highly sensitive to background noise. Collectively, these studies underscore the need for more adaptable segmentation frameworks that can preserve spatial relationships in dense handwritten layouts such as medical prescriptions.

2.2. Other advanced deep learning models in text-line segmentation

Other than U-Net model, various deep learning models like Mask R-CNN, generative adversarial network (GAN), and fully convolutional network (FCN) models showcased their efficacy in segmenting text-lines from various types of documents. To identify text

lines in historical texts, Jian et al. [13] proposed an Iterative Attention Head (IAH) and a Dynamic Rotational Proposal Network (DRPN), that integrates into Mask R-CNN. Dorby et al. [14] worked on extraction of text-lines from by training the document patches using Mask R-CNN and further merged these patches to segment text-lines from the whole page of historical documents. A study by Fizaine et al. [15] directly compares the U-Net-based models with Mask R-CNN and claims outperformance in segmenting text-lines.

A hybrid method introduced by Vo et al. [16] predicts the structure of text-line utilizing an FCN, and generates a line map for text strings and line adjacency graph (LAG) method that splits the touching characters. Another FCN-based U-shaped model, Doc-UFCN, proposed by Biollet et al. [17] for detecting the text-lines from historical documents, emphasizes the usage of a lighter pre-training architecture instead of heavy encoders like ResNet. A novel strategy to extract text-lines using GAN was introduced by Kundu et al. [18]. This approach employs U-Net architecture for generator and Patch GAN for discriminator to address the challenges such as skewed, degraded, and multi-language environments. Along with these challenges, the problem of overlapping characters was also addressed by Demir et al. [19] and Ozseker et al. [20] using GAN, where image-to-image translation was employed to learn the features of text-lines and to generate segmentation masks without post-processing.

2.3. U-Net and its evolution in document and medical image segmentations

The advent of U-Net architectures revolutionized segmentation tasks across multiple domains, particularly in biomedical imaging emphasizing the architecture's flexibility and success in delineating complex anatomical boundaries.

A number of architectural modifications have been proposed to enhance U-Net's performance. For instance, Azad et al. [21] demonstrated the superiority of U-Net variants on diverse medical benchmarks, while Yu et al. [22] proposed EU-Net, an automatically optimized U-Net variant based on evolutionary neural architecture search, further refining segmentation performance. Other notable works include DC-UNet [23], and transformer-augmented hybrids [24], each targeting different limitations in feature representation, model depth, and global context awareness.

The discipline of document layout analysis, has seen the emergence of U-Net adaptations. Mechi et al. [25] developed an adaptive U-Net for text-line segmentation in historical documents, showing its effectiveness in preserving structural coherence. However, most U-Net applications in document analysis remain constrained to line or region-level segmentation, lacking the granularity to preserve block-level semantics.

Although U-Net and its variants have exhibited strong interpretation in both medical and document segmentation tasks, their application to the specific challenge of prescription text-block segmentation remains underexplored. Medical prescriptions require not only accurate boundary detection but also contextual preservation between medicine names and associated instructions. Existing models do not adequately address this nuanced requirement, indicating a clear research gap. The present work responds to this need by proposing an attention-augmented U-Net framework tailored for fine-grained block segmentation in handwritten medical prescriptions.

2.4. Attention mechanisms in document understanding

The incorporation of attention mechanisms into deep learning architectures has significantly improved the interpretability and performance of models across various document understanding tasks [26]. Originally introduced to enhance natural language processing, attention mechanisms have been effectively adapted to visual tasks, including document layout analysis and page object detection. Naik et al. [27] explored the role of attention in detecting structural components within document images, showing that attention-based models could localize and classify layout elements more effectively than conventional CNN-based approaches.

The utility of attention as a tool for interpretability has also been emphasized. Tutek and Snajder [28] investigated the practical deployment of attention mechanisms for explainable artificial intelligence applications, while Soydaner [29] provided a comprehensive analysis of how attention operates within neural networks across various domains. These findings were further reinforced by Brauwers and Frasincar [30], who presented an extensive survey of attention-based models, underscoring their widespread applicability and impact on model performance.

In the context of medical and document image segmentations, mechanisms of attention have been embedded within U-Net variants to enhance localization accuracy. The works such as SA-UNet [31] and ASCU-Net [32], each integrated spatial and channel attention to better capture hierarchical features. The CBAM module [33], combining both types of attention, has proven particularly effective in guiding convolutional networks to emphasize informative features in medical and layout images.

Despite these advancements, most attention-based models in the literature have focused on general image segmentation or high-level document component recognition. Cao et al. [34] proposed selective region concentration for visual document understanding but primarily targeted forms and structured layouts. In the domain of handwritten medical prescriptions, existing works such as those by Hassan et al. [35] and Jain et al. [36] have concentrated on character recognition or end-to-end prescription transcription using CNN-LSTM and CRNN models. While effective for isolated token recognition, these approaches often fail to preserve the spatial–semantic relationships among grouped entities like medicine names, dosages, and administration instructions.

Furthermore, document-specific models such as DocPresRec [37] and attention guided recognition methods for student notes [38] remain largely line-centric or token-based, lacking the granularity required to segment and preserve functional blocks of medical prescriptions. Shende et al. [39] addressed handwriting recognition and prescription scanning, but their method overlooked the need to preserve the hierarchical grouping of textual components.

This body of work reveals a significant gap: while attention mechanisms have enhanced structural understanding in many document processing scenarios, their application to the block-level segmentation of handwritten medical prescriptions remains underdeveloped. Medical prescriptions are inherently spatially dense and semantically interdependent, where the failure to preserve block structure may lead to the disassociation of critical information. The current research addresses this gap by proposing a U-Net-based segmentation framework augmented with spatial and channel attention, explicitly designed to extract and preserve coherent prescription blocks from complex handwritten inputs.

2.5. Motivation and challenges

Table 1 provides the details of various recent deep learning models used in text-line segmentation with architectural highlights and limitations on handwritten documents along with results of different evaluation parameters. Each study focuses on text-line segmentation and none focuses on block-level segmentation. This motivates our study to experiment on text block-level segmentation using handwritten medical prescriptions.

Table 1
Comparison of deep learning models for text-line segmentation

Model	Architectural highlights	Limitations in handwritten documents	Results
UFCN [17]	U-Net with FCN using dilated convolutions	Model trained on one dataset often do not generalize well on other historical dataset as each dataset uses different annotation formats and conventions.	IoU — 0.80, *F*1-score — 0.89, AP@0.5–0.94 on ScribbleLense dataset
Adaptive U-Net [25]	U-Net with 32 filters at the initial block to reduce parameters	Performance drops on very complex/ variable layouts	Precision — 0.75, Recall — 0.85, *F*-score — 0.79 on cBad Dataset
Attention U-Net [40]	U-Net with attention gates for spatial focus	Still may struggle without domain specific tuning in cluttered documents	Precision — 0.93, Recall — 0.94, *F*-measure — 0.93 on BADAM dataset
Vision transformer-based model [41]	Captures global dependencies; learns contextual token relationships	Requires large datasets and longer training; sensitive to irregular cursive flow	Detection rate — 0.92 on Turkish Line segmentation dataset
Mask R-CNN [15]	Uses Region Proposal Network (RPN) for instance segmentation with bounding boxes and masks	Requires precise bounding box annotations and careful hyper-parameter tuning	IoU — 0.85, *F*1-score — 0.91, AP@0.5–0.98 on HOME-Alcar dataset
GAN [20]	Conditional GAN with encoder–decoder generator and learns structured translation from images to masks. Captures global and fine texture context	Heavy model and needs diverse training data. Sensitive to discriminator and generator balance	GAN loss — 0.99, GAN L1 — 0.94, GAN L2 — 0.90 on VML-AHTE dataset

Despite considerable progress in document layout analysis, handwritten text segmentation, and attention-based neural models, several persistent gaps remain unaddressed in the literature—particularly concerning the segmentation of handwritten medical prescriptions. The following research challenges have been identified:

Block-level preservation in dense handwritten layouts: Existing methods such as Jain et al. [36] and Shende et al. [39] focus primarily on line or character-level segmentation and recognition, often overlooking the need for block level semantic coherence in handwritten medical prescriptions. This omission may lead to disassociation of critical information such as drug names and corresponding dosage instructions.

Limited adaptability of traditional deep segmentation models: Conventional segmentation models including FCN, SegNet, and even standard U-Net variants often struggle to maintain high fidelity in scenarios involving spatial noise, overlapping strokes, and low-contrast ink conditions. Work by Shivakumara et al. [12] demonstrates this vulnerability when applied to real-world clinical data.

Insufficient integration of attention in fine-grained document tasks: Although attention mechanisms have been widely adopted for high-level document classification and structural component detection [27], their application in enhancing block-level segmentation in handwritten contexts remains underexplored. Notable frameworks such as Attention U-Net and CBAM [33] have shown promise in medical imaging but are seldom optimized for unstructured, densely packed handwritten prescriptions.

Scarcity of benchmarked frameworks for medical prescription layout segmentation: End-to-end systems such as Seamformer [11] and DocPresRec [37] target holistic understanding or structured layout inference but lack dedicated mechanisms for preserving functional groupings at a block level. Most available systems do not benchmark their segmentation modules independently for medical prescriptions, thereby limiting generalizability assessments.

These gaps motivate and provide a baseline for the present study to introduce a novel attention-augmented U-Net architecture explicitly tailored to handwritten medical prescriptions. This work prioritizes block-level integrity of the medicines and its components to preserve the association between them and integrates spatial–channel attention. This comprehensive approach positions the proposed method as a distinct advancement over existing segmentation pipelines in prescription digitization.

3. Proposed Methodology

The objective of the proposed framework is to segment structured textual blocks from handwritten medical prescriptions. Formally, let $X \in R^{H\times W}$ denote a grayscale input image and $Y \in \{0,1\}^{H\times W}$ its corresponding ground truth binary mask, where each pixel $y_{i,j}$ indicates whether pixel (i,j) belongs to a meaningful text block grouping (e.g., medicine name, dosage, and frequency).

The segmentation task is modeled as a dense binary classification setback where the intent is to learn a mapping function $f_\theta: R^{H\times W} \to \{0,1\}^{H\times W}$ parameterized by network weights θ, such that:

$$\widehat{Y} = f_\theta(X) = \sigma(W * \phi(X) + b) \tag{1}$$

where $\phi(\cdot)$ denotes the encoder–decoder network (i.e., U-Net with attention), W implies a 1 × 1 convolutional projection layer, and $\sigma(\cdot)$ is the sigmoid activation function that outputs pixel-wise probabilities in [0,1]. Each prediction $\widehat{y_{l,j}} > 0.5$ is interpreted as a foreground pixel, contributing to a coherent prescription block.

Unlike line-level segmentation techniques, this formulation aims to preserve the spatial and semantic integrity of block-level groupings, mitigating the risk of disassociating interdependent prescription elements (e.g., breaking dosage from its corresponding drug name). The model is thus optimized to generate masks $\widehat{Y} \in \{0,1\}^{H\times W}$ indicating the spatial footprint of prescription blocks that maximize both pixel-wise accuracy and region-wise cohesion, which are measured using a composite loss $\mathcal{L}$ defined as:

$$\mathcal{L}(\hat{Y}, Y) = \alpha \cdot \mathrm{BCE}(\hat{Y}, Y) + (1 - \alpha) \cdot \mathrm{Dice}(\hat{Y}, Y) \tag{2}$$

where $\alpha \in [0,1]$ balances the binary cross-entropy (BCE) and Dice loss terms.

The Dice measures the region-overlap-based metric and evaluates how well the predicted mask ($\widehat{Y}$) overlaps with the ground-truth mask (Y).

$$\mathrm{Dice}\left(\widehat{Y}, Y\right) = \frac{2\sum_{i,j} \widehat{y_{l,j}} \cdot y_{i,j}}{\sum_{i,j} \widehat{y_{l,j}} + \sum_{i,j} y_{i,j}} \tag{3}$$

where $\widehat{y_{l,j}}$ is the predicted probability at pixel (i,j), and $y_{i,j}$ is the corresponding ground-truth. Dice emphasizes block-wise segmentation where foreground pixels are sparse and spatial cohesion is demanding. Here, the Dice is integrated as a loss term to encourage coherent block-level segmentation.

The end-to-end segmentation pipeline is schematically illustrated in Figure 2. The ROI extracted image along with its masks are passed as input to the proposed model. The masks are generated using the annotation tools that serves as a ground-truth during the training phase.

3.1. Encoder–Decoder backbone with Attention modulation

The core architecture employed in this study is a refined variant of the canonical U-Net, herein referred to as PrescNet. This design integrates three principal innovations: (i) an auxiliary shallow convolutional stem for early feature enhancement, (ii) attention gating across skip connections to enable discriminative feature selection, and (iii) combined spatial and channel attention mechanisms to reinforce semantic localization. These modifications are engineered specifically for the task of block-wise segmentation of handwritten prescriptions, which are characterized by spatial clutter, overlapping strokes, and cursive variability.

The encoder path comprises five hierarchical stages. At each level l, the input tensor F_l is transformed via two successive convolutional operations, each defined as:

$$F_l' = \phi\left(B\left(W_l^{(2)} * \phi\left(B\left(W_l^{(1)} * F_l\right)\right)\right)\right) \tag{4}$$

where $W_l^{(1)}$ and $W_l^{(2)}$ denote the 3 × 3 convolutional kernels at level l, $B(\cdot)$ symbolizes batch normalization, and $\phi(\cdot)$ denotes the ReLU activation function.

The batch normalization, applied to the output of each convolution layer normalizes feature activations computed using the mean and variance over mini-batch and applies learnable scaling and shifting parameters. For a given activation 'x', batch normalization is defined as,

$$B(x) = \gamma \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta \tag{5}$$

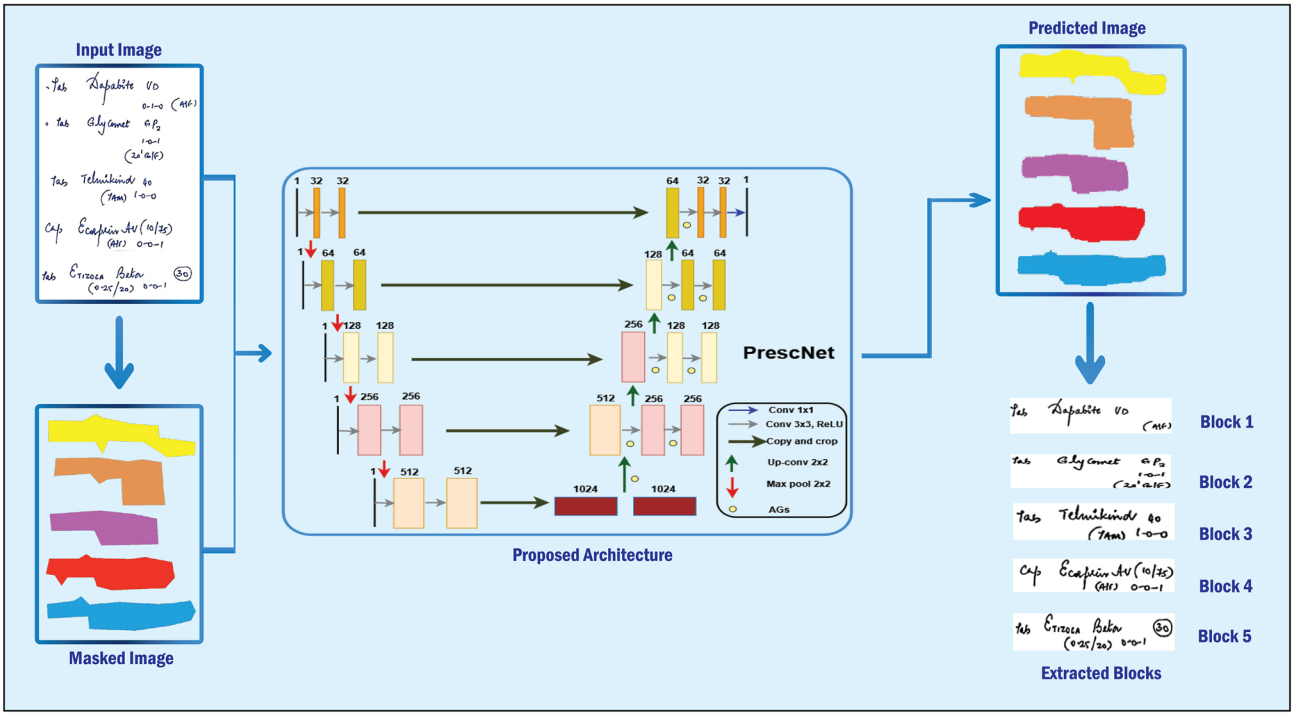
where μ and σ^2 are the mean and variance of batch and γ and β are the learnable parameters. $B(\cdot)$ generates real-value feature maps with stabilized distribution, that improves training stability by accelerating convergence and reduces sensitivity to handwriting variability occur in prescription images.

A shallow 32-channel projection layer is introduced as the initial convolutional block $P(\cdot)$ operating on the input X:

$$F_0 = \phi(B(W_P * X + b_P)) \tag{6}$$

This layer serves as a low-level feature amplifier that progressively down-samples the input image designed to capture micro-textural cues that may otherwise be attenuated by deeper layers, a critical factor for enhancing character boundary localization in cluttered scripts. The decoder reverses the encoder structure by

Figure 2
Proposed PrescNet model



applying transposed convolutions to progressively restore spatial resolution and produce segmentation maps. Importantly, skip connections between encoder and decoder blocks are modulated by attention gates, as opposed to naive concatenation.

3.1.1. Attention-gated skip connections

The attention gated skip connections preserve the fine structural details that are lost during down-sampling and transfer the intermediate encoder features to the decoder by enabling precise boundary delineations.

Let g_l represent the gating signal from decoder level l and x_l be the corresponding encoder output. The attention gating function $A(\cdot,\cdot)$ is computed as:

$$\psi_l = \sigma(W_\psi \cdot \phi(W_g g_l + W_x x_l) + b) \tag{7}$$

where $\sigma(\cdot)$ is the sigmoid function and ψ_l denotes the spatial attention coefficients for skip connection l. The filtered feature map is defined as:

$$x_l' = \psi_l \odot x_l \tag{8}$$

with $\odot$ indicating element-wise multiplication. These modulated features are concatenated with upsampled decoder outputs prior to further decoding.

3.1.2. Channel–spatial attention fusion

The skip connections pass all the features, including noise and some irrelevant background information. Hence, to filter encoder outputs, and to enhance contextual awareness, each decoder feature map undergoes channel–spatial attention fusion. These attentions emphasize on the most discriminative features and relevant pixel regions, suppressing noise and blank areas and contribute reconstructions. It also helps in improving the separation of overlapping lines or closely spaced lines by enhancing the clarity of segmentation mask by reducing the influence of artifacts. A channel attention map $M_c \in R^{C\times1\times1}$ and a spatial attention map $M_s \in R^{1\times H\times W}$ are computed sequentially and multiplied with the intermediate feature representation F:

$$F' = F \odot M_c \odot M_s \tag{9}$$

These attention maps are learned implicitly and highlight the most discriminative regions and feature channels across the segmentation hierarchy [33].

The integration of shallow feature extraction, spatial attention, and selective skip connections is critical for precise delineation of text boundaries in handwritten prescriptions. Unlike natural image segmentation tasks, the target domain in this work involves low inter-class variance and substantial intra-class deformation. Hence, enhancing intra-layer focus via channel–spatial modulation and inter-layer flow control via attention gates allows for improved generalization under variable scan quality and handwriting styles.

This architectural framework thus obtains the balance between computational efficiency and contextual adaptively, laying the foundation for robust clinical document parsing and downstream information extraction.

3.2. Output prediction and sigmoid projection

The final decoder output is projected to a single-channel mask via a 1 × 1 convolution and sigmoid activation:

$$\widehat{Y} = \sigma(W_{final} \quad * \quad F_{out} + b_{final}) \tag{10}$$

where $\widehat{Y}$ represents the probabilistic segmentation mask with pixel-wise membership scores in [0,1]. A threshold of 0.5 is applied during inference to obtain the binary label map.

The textual contents from the individual masks are extracted from the output predicted masks as text-blocks for our future study.

3.3. Loss function and optimization strategy

To effectively train the segmentation model in the presence of class imbalance and noisy annotations, a hybrid loss function is employed by linearly uniting the BCE loss and the Dice loss. This composite objective function offers a principled balance between pixel-wise classification accuracy and regional overlap fidelity, which is crucial for dense text-block segmentation where boundary delineation is inherently ambiguous.

Let $\widehat{Y} \in [0,1]^{H\times W}$ denote the predicted segmentation mask and $\widehat{Y} \in \{0,1\}^{H\times W}$ the corresponding ground truth binary mask. The BCE loss $\mathcal{L}_{\text{BCE}}$ is defined as:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log\left(\widehat{y_i}\right) + (1-y_i)\log\left(1-\widehat{y_i}\right)\right] \tag{11}$$

where $N = H \times W$ is the total number of pixels, and y_i and $\widehat{y_i}$ represent the ground truth and predicted values at pixel i, respectively. While BCE penalizes incorrect pixel classifications, it treats all pixels equally and is sensitive to class imbalance.

To address these hindrances, the Dice loss $\mathcal{L}_{\text{Dice}}$ is introduced, which estimates the overlap between ground truth and prediction:

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_{i=1}^{N} y_i \widehat{y}_i + \epsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \widehat{y}_i + \epsilon}, \tag{12}$$

where ϵ is a smoothing constant $\left(\epsilon = 10^{-6}\right)$ to ensure numerical stability. The Dice loss emphasizes structural similarity and penalizes under-segmentation more heavily than pixel misclassification.

The net combined loss $\mathcal{L}_{\text{total}}$ is a weighted sum of both terms:

$$\mathcal{L}_{total} = \alpha \cdot \mathcal{L}_{\text{BCE}} + (1-\alpha) \cdot \mathcal{L}_{\text{Dice}} \tag{13}$$

where $\alpha \in [0,1]$ controls the trade-off between pixel-wise accuracy and region level consistency. In our experiments, α was empirically set to 0.5 to ensure equal importance during optimization. This dual-objective formulation has been shown to stabilize training while simultaneously improving segmentation robustness in small and imbalanced foreground regions [21].

Dice coefficient (D): evaluates spatial overlap, equivalent to $F1$-score in binary tasks,

$$D = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{14}$$

Intersection over Union (IoU): assesses normalized region agreement between prediction $\widehat{Y}$ and ground truth Y,

$$IoU = \frac{/\widehat{Y} \cap Y/}{/\widehat{Y} \cup Y/} \tag{15}$$

Precision (P) and Recall (R): Quantify pixel-level correctness and completeness,

$$P = \frac{TP}{TP + FP} \text{ and } R = \frac{TP}{TP + FN} \tag{16}$$

These metrics altogether provides nuanced understanding of the model performance, capturing both micro-level pixel fidelity and macro-level boundary coherence—critical in clinical document scenarios where precise text-block localization directly influences downstream interpretation and digitization.

4. Experimentation and Results

To rigorously examine the efficacy of the proposed PrescNet architecture, a series of controlled experiments were conducted. This section presents a dataset description, comparative analysis of different model variants, training–validation split ratios, and the associated performance across multiple segmentation metrics. The goal is to verify the superiority of the proposed method through empirical evidence and establish its robustness across configurations.

4.1. Dataset description

The efficacy of deep learning models in prescription segmentation is closely tied to the complexity and variability present in the training data. For this study, we employ a custom dataset comprising 855 grayscale image–mask pairs derived from handwritten medical prescriptions. Unlike public datasets, the current dataset was curated through an automatic extraction process of ROI/advice section from the actual prescription [42] and underwent post-processing procedures as outlined in earlier internal works. Each prescription image was manually labeled to produce corresponding binary segmentation masks that delineate entire semantic blocks rather than isolated words or characters. A block here denotes a contiguous region comprising elements such as drug name, dosage, and administration frequency. The goal is to capture such medically meaningful groupings, which form the basis for accurate digitization and downstream extraction.

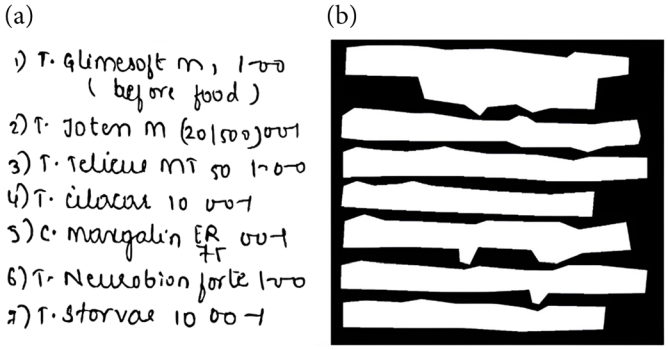
The dataset is sliced into different training and validation images, maintaining some partition and the images are rescaled to 256 × 256 and normalized to have zero mean and unit variance. To improve model generalization over diverse handwriting styles, random horizontal flips and contrast adjustments are applied as part of data augmentation. The transformation pipeline T applied to each sample X is defined as:

$$T(X) = \text{Norm} \circ \text{Flipp} = 0.5 \circ \text{Resize}256 \times 256(X) \quad (17)$$

where Norm(·) denotes standard normalization and Flipp = 0.5 refers to stochastic horizontal flipping with a probability of 0.5.

Figure 3 presents several representative samples from the dataset. The left picture shows the raw grayscale prescription image, while the right displays the annotated segmentation mask. These masks are block-level in nature, encapsulating semantically grouped handwritten content and captures entire semantic units essential for structured interpretation. This block-level annotation facilitates downstream tasks such as drug extraction, dosage parsing, and entity linking, forming a critical part of the prescription digitization pipeline.

Figure 3
Representative sample from the dataset: (a) original grayscale prescription and (b) manually annotated binary mask capturing contiguous text blocks



4.2. Training setup and evaluation protocol

The model training was conducted by adopting PyTorch v1.13.1 on NVIDIA-Tesla T4 GPU of 16 GB memory. To maintain computational efficiency and reproducibility, the proposed PrescNet U-Net was trained for 50 epochs, which yielded a validation accuracy of 98%. The 50-epoch configuration was selected for all reported experiments to balance training time and overfitting risks. This choice aligns with empirical observations across multiple folds, where performance saturated within the 45–50 epoch range, indicating sufficient convergence in order to scale the dataset.

A batch size of 4 was chosen after empirical benchmarking to ensure optimal memory utilization without compromising gradient stability. Given the resolution and complexity of handwritten prescription images, larger bath sizes led to memory exhaustion on the available GPU, whereas smaller batches induced unstable updates due to high variance in mini-batch gradients. The selection of batch size = 4 thus represents a compromise between stable convergence and hardware constraints, particularly effective for medium-sized medical imaging datasets like the one which is used.

The dataset was divided into nine distinct training–validation ratios: 90:10, 80:20, 70:30, 60:40, 50:50, 40:60, 30:70, 20:80, and 10:90 with stratified random sampling used to preserve label balance across partitions. The 70:30 split was found to provide the best generalization performance, offering a reliable trade-off between model robustness and training data sufficiency. Consequently, all primary experiments and performance evaluations were conducted using this partition.

To statistically evaluate generalization and mitigate sampling variance, a 10-fold cross-validation protocol was utilized. The full dataset D was divided into 10 mutually exclusive folds $\{D1,...,D10\}$, such that each fold served once as a validation set while the remaining nine formed the training subset. This strategy produced averaged performance metrics and reduced dependency on any single data configuration, thereby increasing result reliability.

Training optimization employed the Adam algorithm with learning rate of 10^{-4} and no scheduler, allowing analysis of the model's inherent generalization capacity without external learning rate modulation. All experiments used identical hyperparameter settings to ensure fair comparisons across folds and data splits.

Model evaluation was performed using a comprehensive set of metrics for binary segmentation, encompassing both pixel-level and region-level assessments.

4.3. Data split evaluation

To understand the impact of dataset partitioning on model generalization, we conducted controlled experiments using nine training–validation splits: 90:10, 80:20, 70:30, 60:40, 50:50, 40:60, 30:70, 20:80, and 10:90. The same model architecture (PrescNet) and hyperparameters were maintained across all configurations to ensure consistency. Table 2 summarizes the averaged results, showcasing key evaluation metrics including validation IoU, validation Dice score, and validation Dice loss.

The val IoU and val Dice scores show optimal performance for the 70:30 split. This observation highlights a key insight: increasing the validation set size can result in reduced exposure to training variability, thus undermining the model's ability to generalize spatial boundaries.

Table 2
Performance comparison across different train–validation splits

Split ratio	Val IoU	Val Dice score	Val Dice loss	Val precision	Val recall
90:10	86.09	92.22	0.078	93.86	91.40
80:20	86.13	92.24	0.076	**94.42**	90.91
70:30	**87.21**	**92.91**	**0.071**	92.56	**93.91**
60:40	86.18	92.22	0.078	93.92	91.45
50:50	85.95	92.08	0.079	92.80	92.26
40:60	85.77	91.81	0.082	92.27	92.6
30:70	85.24	91.38	0.086	93.56	90.77
20:80	85.07	90.46	0.095	91.68	92.41
10:90	81.22	85.33	0.147	94.41	85.56

Consequently, the 70:30 partition is chosen as the primary configuration in all downstream evaluations. Dice loss, which quantifies the overlap sensitivity maintains dominance in minimizing segmentation boundary mismatches. With the same 70:30 train–val split ratio, the proposed model achieved a val precision of 92.56% and val recall of 93.91%, which implies a strong performance in correctly identifying relevant blocks while lowering the false negatives and false positives. Besides, the model reported a BCE loss of 0.258, reflecting a good alignment between actual and predicted pixel-wise segmentations. The effective optimization and generalization on the validation set are demonstrated with the combined loss, that was reduced to 0.167. Collectively, the analysis across each of the data splits reinforces the empirical decision to adopt the 70:30 split for all model variants and different deep learning models to accomplish comparative experiments due to its optimal trade-off between loss minimization and validation stability.

4.4. Ablation study

This section systematically evaluates the effectiveness of incorporating lightweight 32 channel projection along with the spatial–channel attention gates to classical U-Net model. Table 3 summarizes the ablation study conducted to analyze the contribution of different architectural components. The vanilla U-Net establishes a baseline performance by achieving a validation IoU of 83.78% and Dice score of 90.62%. The U-Net with 32 channels projection has showed better results in terms of validation metrics compared to baseline. Further, the spatial–channel attention gates are incorporated to encoder section of the U-Net, and observed marginally improved metrics indicates sharper segmentation. The vanilla U-Net incorporated with attentions in decoder as well as in encoder and decoder attained better performance than baseline but more or less same performance with lightweight channels. CBAM based U-Net also improved upon the baseline but lagged behind the tailored attention mechanism. Additionally, it is observed that constituting attention gates either in encoder or decoder or in both have almost the same performance.

Table 3
Performance comparison across attention and channel in U-Net model architectures

Model	Val IoU	Val Dice score	Val Dice loss	Val precision	Val recall
U-Net	83.78	90.62	0.094	91.5	91.04
U-Net + 32 channels	85.8	91.1	0.080	92.08	90.91
U-Net + 32 channels + Attention (E)	86.21	92	0.077	93.19	92.11
U-Net + Attention (D)	84.62	91.21	0.087	92.17	89.24
U-Net + Attention (E & D)	85.92	91.70	0.082	**93.61**	91.45
U-Net + 32 channels+ Attention (E & D)	86.20	92	0.077	92.5	92.8
U-Net + 32 channels + CBAM	85.51	91.32	0.087	92.42	92
PrescNet	**87.21**	**92.90**	**0.071**	92.56	**93.91**

4.5. Quantitative analysis

Our proposed PrescNet model achieves the best overall performance, with a validation IoU of 87.21% and Dice score of 92.9% with the lowest Dice loss of 0.071. Notably, it attained better precision of 92.56% and highest recall of 93.91%, outperforming all ablations. These performance results confirm that the integration of channel–spatial attention method with lightweight channel model emphasizes discriminative text-stroke patterns while suppressing the background clutter.

To validate the contributions of the architectural modification and to determine the potency of the proposed model, we carried out a comparative analysis against leading-edge deep learning models and few variants of U-Net models such as, CNN, FCN, UFCN, GANPatch, Mask R-CNN, U-Net, Attention U-Net, CBAM U-Net, Adapt U-Net, and the final PrescNet model. All models were trained under the same 70:30 train–validation split, using identical hyperparameters and optimization protocols. The evaluation was conducted using key segmentation metrics: validation IoU, validation Dice coefficient, and validation Dice loss.

Table 4 exhibits an analytical comparison of the proposed PrescNet model relative to few established deep models for the task of text-block semantic segmentation in handwritten medical prescriptions. The evaluation encompasses validation metrics and loss functions are reported.

The proposed PrescNet model demonstrates superior performance across all the evaluation metrics and loss. Specifically, it achieves the highest IoU (87.21%), Dice score (92.9%), and optimal Dice loss (0.071), and improved generalization on validation data indicates a

Table 4
Performance comparison across different deep-learning model architectures

Model	Val IoU	Val Dice score	Val Dice loss	Val precision	Val recall
CNN	79.40	88.39	0.163	89.7	87.86
FCN	80.59	88.01	0.124	85.44	91.05
UFCN	76.22	86.47	0.13	91.31	82.35
GANPatch	77.38	87.22	0.122	89.92	84.89
Mask R-CNN	84.56	91.58	0.414	87.89	**94.84**
U-Net	83.78	90.62	0.094	91.5	91.04
Adapt U-Net	83.15	90.05	0.10	90.79	91.5
PrescNet	**87.21**	**92.90**	**0.071**	**92.56**	93.91

more accurate overlap between the ground-truth and predicted masks. Further, it also yields a good precision (92.56%) and recall (93.91%), highlighting the robustness in identifying true positive regions and low false detections. This indicates that the integration of attention gates and initial 32-channel projections enhances the segmentation precision, particularly for irregular and cluttered handwriting instances.

Among the existing methods, Mask R-CNN shows relatively competitive performance with IoU of 84.56% and Dice score of 91.58%. The standard U-Net also performs well with IoU of 83.78% and Dice score of 90.62%, confirming its effectiveness as a robust baseline for segmentation task. In contrast, the conventional models such as CNN and FCN achieve lower performance while UFCN and GAN exhibit even weaker results with IoU and Dice score.

4.6. 10-fold cross-validation and error analysis

4.6.1. Fold-wise metric trajectories

To further validate consistency across training folds, Figure 4(a)–(e) illustrate the metric progression across 10 folds. These plots reflect how performance evolved across training epochs (50 total), highlighting the learning stability and convergence behavior of the PrescNet. However, the instance of fold 7 provides the best result than the rest of the folds with respect to validation IoU, Dice score, and Dice loss.

The IoU metric, visualized in Figure 4(a), remains tightly bounded with no major fold-specific degradation. This supports the high overlap between predicted and ground-truth mask regions. Figure 4(b) indicates that even under varied validation subsets, the model achieves consistent validation Dice gains, reaffirming segmentation metrics across fold-level handwriting diversity.

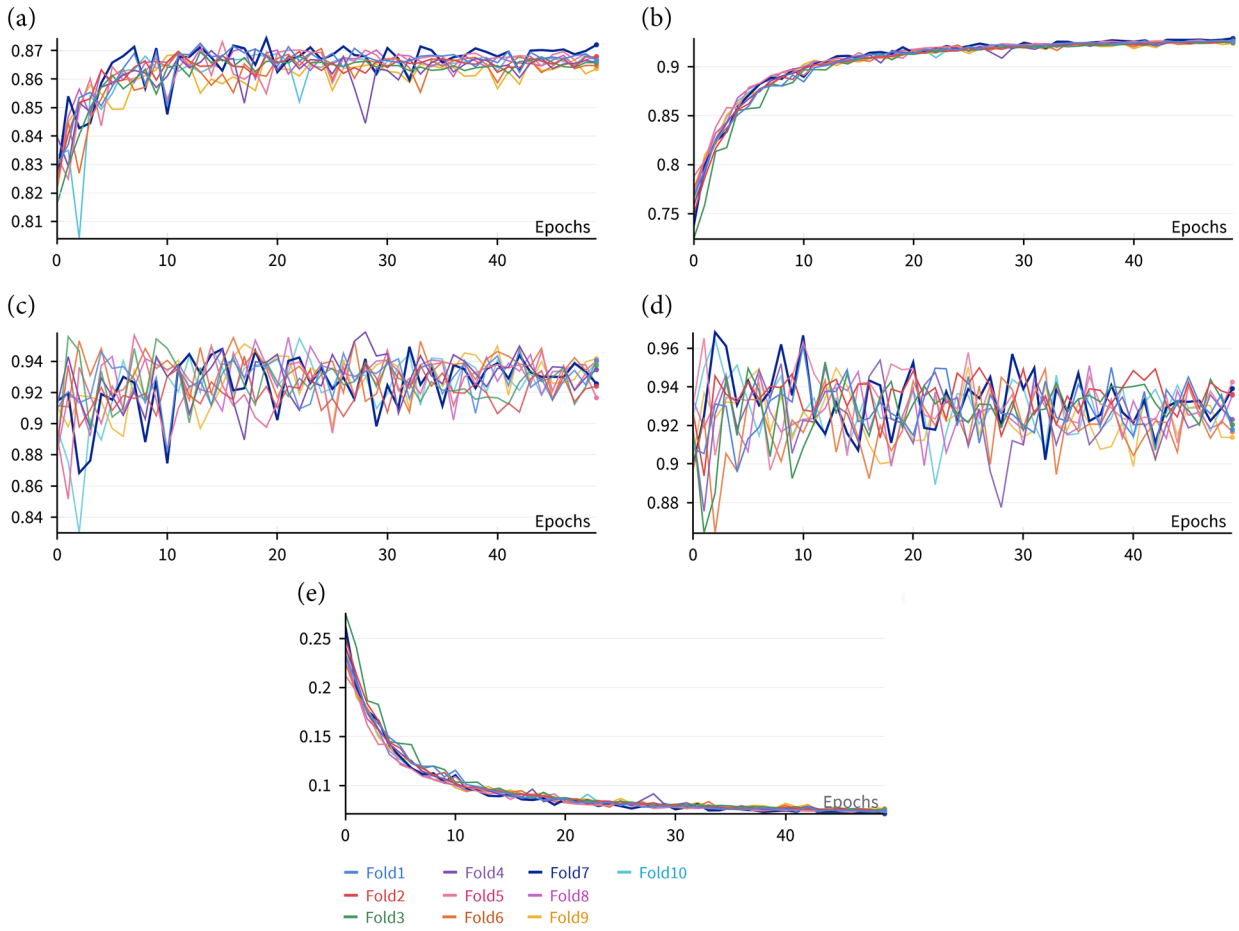
Figure 4(c) presents the fold-wise growth of precision. Despite fold-wise initialization randomness, the final convergence levels are tightly grouped, affirming boundary sensitivity in true positive predictions. As seen in Figure 4(d), recall values gradually reach above 0.90, plateauing with fold-wise divergence. This implies high model sensitivity across varied stroke densities and writing pressure scenarios. In Figure 4(e), the Dice loss drops steadily across all folds, suggesting uniform learning of spatial segmentation fidelity. The minor deviations observed in folds 3, 4, and 6 do not significantly impact downstream performance.

The loss trends are consistent across all folds, exhibit cross-validation splits, are balanced, and the smooth convergence across all folds attests to the architectural stability of PrescNet and confirms that the model maintains strong pixel-level and region-level coherence. These insights reinforce the statistical robustness of the proposed approach and strengthen confidence in its deployment readiness for real-world prescription digitization scenarios.

4.6.2. Qualitative error analysis

While the PrescNet demonstrates commendable results in terms of IoU, Dice score, and generalizability, certain challenging cases

Figure 4
Fold-wise metric trajectories across 10-fold cross-validation



highlight residual limitations, particularly under conditions such as ink fading, text overlap, and a typical spatial layouts. A focused qualitative error analysis was conducted on complex samples from the test set. Figures 5 and 6 present two representative examples that reveal characteristic error modes observed in real-world handwritten prescriptions.

Figure 5 displays a vertically stacked prescription with slanted cursive handwriting. Although the network delineates four out of five textual segments, the first two blocks are erroneously merged due to overlapping curvature and proximity. This mis-segmentation underscores the challenge of preserving spatial separation in non-linear writing trajectories where adjacent baselines are poorly defined.

As illustrated in Figure 6, a slight under-segmentation errors are observed in lines 2, 4, and 6 caused by adjacent lines. Specifically, line 2, which actually belongs to block1, is incorrectly merged with block2. Furthermore, the second and third blocks have been erroneously fragmented into small sub-blocks signifying incorrect segmentation. This indicates the need for regularization that distinguish complete block boundaries.

Figure 5
Prediction for a vertically aligned, sparse prescription layout (sample 1)

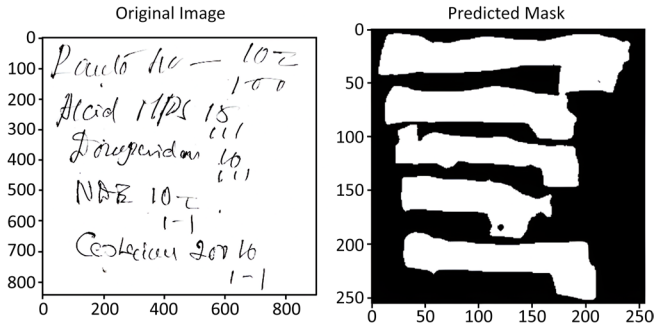
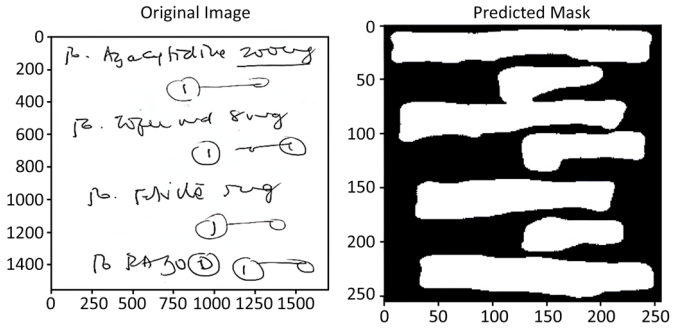


Figure 6
Prediction for a cluttered prescription (sample 2)



5. Discussion

The superior performance of the proposed PrescNet stems from its deliberate architectural enhancements and context-aware design philosophy. Unlike conventional segmentation networks, the proposed model employs spatial–channel attention mechanisms and gated skip connections that enable the network to selectively prioritize salient features while suppressing background noise and inter-line interference.

A key contributing factor to performance improvement is the integration of attention gates at each decoder–encoder junction, which helps the network dynamically refine feature propagation during up-sampling. This selective filtering ensures that only the most contextually relevant activations are forwarded, thereby improving text-block delineation, especially in cluttered prescriptions. Additionally, the inclusion of a 32-channel projection layer in the encoder promotes better spatial granularity during early-stage feature extraction. As observed in Section 4, this architectural refinement leads to notable gains in both Dice coefficient and IoU across all validation folds.

Block-wise segmentation is particularly critical in the context of handwritten medical prescriptions, where structured layout often encodes implicit semantics—such as medication type, medication names, dosages, and intake frequency. Unlike sentence-based OCR, prescription parsing necessitates reliable detection of spatial groupings to avoid incorrect interpretations or skipped annotations. Therefore, block-level segmentation serves as a vital preprocessing stage for downstream tasks such as named-entity recognition, dosage parsing, and automated e-prescription generation.

The PrescNet demonstrates strong generalization across multiple data split variants (90:10, 80:20, 70:30, 60:40, 50:50, 40:60, 30:70, 20:80, and 10:90) and 10-fold cross-validation settings. Minimal variance in validation IoU, Dice scores, and precision metrics indicates that the model does not overfit to any specific writing style or layout template. Such robustness is critical for real-world deployment, where unseen prescriptions can vary widely in format and clarity.

Despite its strengths, the model exhibits limitations under certain pathological conditions. Handwritten scripts with uncommon flourishes, tight line spacing, or dense overlapping strokes can degrade prediction quality. As illustrated in Figures 5 and 6, these anomalies lead to either under-segmentation or misclassification at boundary regions. Additionally, rare writing patterns such as extremely slanted text or hybrid cursive–print styles challenge the model's learned priors and require further tuning or augmentation.

Future work aims to address these limitations by incorporating multi-view ensemble techniques and temporal learning for layout prediction consistency. Integration of unsupervised spatial priors and reinforcement-driven segmentation policies is also under consideration to enhance performance in visually ambiguous scenarios. Overall, the PrescNet provides a robust, interpretable, and extensible foundation for handwritten medical document analysis in real-world settings.

6. Conclusion

The study outlines an innovative deep learning architecture for the robust segmentation of handwritten medical prescriptions at the block level instead of segmentation at line level to preserve the association between the treatment regimen (medicine and its components). By embedding attention gates into the classical U-Net framework and introducing a shallow 32-channel projection layer, the proposed model effectively captures spatial and contextual dependencies inherent in complex handwritten inputs. Extensive experiments conducted on a custom-curated dataset of prescription images as no dataset of prescriptions are publicly available, demonstrate that the architecture not only enhances segmentation of IoU and Dice score but also achieves superior generalization across various data splits and cross-validation settings. The integration of spatial and channel-wise attention mechanisms enables the network to emphasize semantically relevant regions while minimizing background noise and stroke-level interference—key challenges in prescription interpretation. Comparative analysis with sophisticated models such as CNN, FCN, Mask R-CNN, UFCN, U-Net, AdaptU-Net, GAN, and variants of U-Net further underscores the architectural efficacy of the proposed design. Beyond empirical performance, this work emphasizes the necessity of block-wise segmentation as a foundational step in

automated medical document analysis. The accurate delineation of textual blocks plays a pivotal role in supporting downstream tasks such as drug name recognition, dosage interpretation, and compliance verification in healthcare settings. However, the study also identifies limitations arising from rare handwriting styles, occluded strokes, and severe inter-line overlaps, which can affect segmentation granularity. Future research will aim to address these challenges through more sophisticated augmentation pipelines, hybrid architectures incorporating transformer modules, and unsupervised pretraining on large-scale medical corpora. In conclusion, the proposed PrescNet advances the state-of-the-art in handwritten prescription block-level segmentation and facilitates a scalable and interpretable foundation for intelligent document understanding in medical informatics.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

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

Data Availability Statement

Data available on request from the corresponding author upon reasonable request.

Author Contribution Statement

Rekha G. R.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Siddesha S.:** Conceptualization, Validation, Formal analysis, Investigation, Writing – review & editing, Supervision, Project administration. **V. N. Manjunath Aradhya:** Validation, Formal analysis, Investigation, Writing – review & editing, Supervision.

References

[1] Romero Villa, W. S., Gregorini Machuca, F. A., & Copaja Cornejo, R. N. (2023). Mobile application to digitize handwritten patient records in Peruvian public hospitals. In *LACCEI International Multi-Conference for Engineering, Education and Technology*, 1–7. https://doi.org/10.18687/LACCEI2023.1.1.1114

[2] Betadur, D., Somu, G., & Naveen Kumar, P. (2023). Digitization of inpatient medical records using electronic writing pads in a teaching hospital. In *Advances in Communication, Devices and Networking*, 279–287. https://doi.org/10.1007/978-981-19-2004-2_24

[3] Patel, S., Gayakvad, B., Solanki, R., Patel, R., & Khunt, D. (2023). Towards the digitization of healthcare record management. In R. Malviya, S. Sundram, B. Prajapati, & S. K. Singh (Eds.), *Human-machine interface: Making healthcare digital*, 411–447. Wiley. https://doi.org/10.1002/9781394200344.ch16

[4] Ma, M.-W., Gao, X.-S., Zhang, Z.-Y., Shang, S.-Y., Jin, L., Liu, P.-L., ..., & Zong, H. (2023). Extracting laboratory test information from paper-based reports. *BMC Medical Informatics and Decision Making*, *23*(1), 251. https://doi.org/10.1186/s12911-023-02346-6

[5] Shao, Y., Zhou, M., Zhong, Y., Wu, T., Han, H., Han, S., ..., & Zhang, D. (2022). FormLM: Recommending creation ideas for online forms by modelling semantic and structural information. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 8133–8149. https://doi.org/10.18653/v1/2022.emnlp-main.557

[6] Pomares-Quimbaya, A., Kreuzthaler, M., & Schulz, S. (2019). Current approaches to identify sections within clinical narratives from electronic health records: A systematic review. *BMC Medical Research Methodology*, *19*(1), 155. https://doi.org/10.1186/s12874-019-0792-y

[7] Nguyen, T. N., Burie, J. C., Le, T. L., & Schweyer, A.-V. (2025). Text line segmentation approach combining deep learning model and traditional image processing techniques—Application to transliteration of Cham manuscripts. *Multimedia Tools and Applications*, *884*(32), 39143–39169. https://doi.org/10.1007/s11042-025-20602-x

[8] Bouh, M. M., Hossain, F., & Ahmed, A. (2023). A machine learning approach to digitize medical history and archive in a standard format. In *Proceedings of the 9th International Conference on Information and Communication Technologies for Ageing Well and e-Health*, 230–236. https://doi.org/10.5220/0011986400003476

[9] Wahl, F. M., Wong, K. Y., & Casey, R. G. (1982). Block segmentation and text extraction in mixed text/image documents. *Computer Graphics and Image Processing*, *20*(4), 375–390. https://doi.org/10.1016/0146-664X(82)90059-4

[10] Li, Y., Zheng, Y., Doermann, D., & Jaeger, S. (2008). Script-independent text line segmentation in freestyle handwritten documents. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(8), 1313–1329. https://doi.org/10.1109/TPAMI.2007.70792

[11] Vadlamudi, N., Krishna, R., & Sarvadevabhatla, R. K. (2023). SeamFormer: High precision text line segmentation for handwritten documents. In *International Conference on Document Analysis and Recognition*, 313–331. https://doi.org/10.1007/978-3-031-41685-9_20

[12] Shivakumara, P., Jain, T., Pal, U., Surana, N., Antonacopoulos, A., & Lu, T. (2022). Text line segmentation from struck-out handwritten document images. *Expert Systems with Applications*, 210, 118266. https://doi.org/10.1016/j.eswa.2022.118266

[13] Jian, C., Jin, L., Liang, L., & Liu, C. (2023). HisDoc R-CNN: Robust Chinese historical document text line detection with dynamic rotational proposal network and iterative attention head. In *International Conference on Document Analysis and Recognition*, 428–445. https://doi.org/10.1007/978-3-031-41676-7_25

[14] Droby, A., Kurar Barakat, B., Alaasam, R., Madi, B., Rabaev, I., & El-Sana, J. (2022). Text line extraction in historical documents using Mask R-CNN. *Signals*, *3*(3), 535–549. https://doi.org/10.3390/signals3030032

[15] Fizaine, F. C., Bard, P., Paindavoine, M., Robin, C., Bouyé, E., Lefèvre, R., & Vinter, A. (2024). Historical text line segmentation using deep learning algorithms: Mask-RCNN against U-Net networks. *Journal of Imaging*, *10*(3), 65. https://doi.org/10.3390/jimaging10030065

[16] Vo, Q. N., Kim, S. H., Yang, H. J., & Lee, G. S. (2018). Text line segmentation using a fully convolutional network in handwritten document images. *IET Image Processing*, *12*(3), 438–446. https://doi.org/10.1049/iet-ipr.2017.0083

[17] Boillet, M., Kermorvant, C., & Paquet, T. (2022). Robust text line detection in historical documents: Learning and evaluation methods. *International Journal on Document Analysis and Recognition*, *25*(2), 95–114. https://doi.org/10.1007/s10032-022-00395-7

[18] Kundu, S., Paul, S., Bera, S. K., Abraham, A., & Sarkar, R. (2020). Text-line extraction from handwritten document images using GAN. *Expert Systems with Applications*, *140*, 112916. https://doi.org/10.1016/j.eswa.2019.112916

[19] Demır, A. A., ÖzŞeker, İ., & Özkaya, U. (2021). Text line segmentation in handwritten documents with generative adversarial networks. In *International Conference on INnovations in Intelligent SysTems and Applications*, 1–5. https://doi.org/10.1109/INISTA52262.2021.9548523

[20] Özşeker, İ., Demir, A. A., & Özkaya, U. (2025). GAN-based text line segmentation method for challenging handwritten documents. *International Journal on Document Analysis and Recognition*, *28*(1), 59–69. https://doi.org/10.1007/s10032-024-00488-5

[21] Azad, R., Aghdam, E. K., Rauland, A., Jia, Y., Avval, A. H., Bozorgpour, A., ..., & Merhof, D. (2024). Medical image segmentation review: The success of U-Net. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *46*(12), 10076–10095. https://doi.org/10.1109/TPAMI.2024.3435571

[22] Yu, C., Wang, Y., Tang, C., Feng, W., & Lv, J. (2023). EU-Net: Automatic U-Net neural architecture search with differential evolutionary algorithm for medical image segmentation. *Computers in Biology and Medicine*, *167*, 107579. https://doi.org/10.1016/j.compbiomed.2023.107579

[23] Lou, A., Guan, S., & Loew, M. H. (2021). DC-UNet: Rethinking the U-Net architecture with dual channel efficient CNN for medical image segmentation. In *Medical Imaging 2021: Image Processing: Proceedings of SPIE*, *11596*, 115962T. https://doi.org/10.1117/12.2582338

[24] Jia, X., Bartlett, J., Zhang, T., Lu, W., Qiu, Z., & Duan, J. (2022). U-Net vs transformer: Is U-Net outdated in medical image registration? In C. Lian, X. Cao, I. Rekik, X. Xu, & Z. Cui (Eds.), *Machine learning in medical imaging*, 151–160. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-21014-3_16

[25] Mechi, O., Mehri, M., Ingold, R., & Essoukri Ben Amara, N. (2019). Text line segmentation in historical document images using an adaptive U-Net architecture. In *International Conference on Document Analysis and Recognition*, 369–374. https://doi.org/10.1109/ICDAR.2019.00066

[26] Chauhan, R., Karnati, M., & Singh, P. (2024). Attention based deep neural network for classification of kidney ailments using CT images. In *International Conference on Computing Communication and Networking Technologies*, 1–5. https://doi.org/10.1109/ICCCNT61001.2024.10723909

[27] Naik, S., Hashmi, K. A., Pagani, A., Liwicki, M., Stricker, D., & Afzal, M. Z. (2022). Investigating attention mechanism for page object detection in document images. *Applied Sciences*, *12*(15), 7486. https://doi.org/10.3390/app12157486

[28] Tutek, M., & Šnajder, J. (2022). Toward practical usage of the attention mechanism as a tool for interpretability. *IEEE Access*, *10*, 47011–47030. https://doi.org/10.1109/ACCESS.2022.3169772

[29] Soydaner, D. (2022). Attention mechanism in neural networks: Where it comes and where it goes. *Neural Computing and Applications*, *34*(16), 13371–13385. https://doi.org/10.1007/s00521-022-07366-3

[30] Brauwers, G., & Frasincar, F. (2023). A general survey on attention mechanisms in deep learning. *IEEE Transactions on Knowledge and Data Engineering*, *35*(4), 3279–3298. https://doi.org/10.1109/TKDE.2021.3126456

[31] Guo, C., Szemenyei, M., Yi, Y., Wang, W., Chen, B., & Fan, C. (2021). SA-Unet: Spatial attention U-Net for retinal vessel segmentation. In *International Conference on Pattern Recognition*, 1236–1242. https://doi.org/10.1109/ICPR48806.2021.9413346

[32] Tong, X., Wei, J., Sun, B., Su, S., Zuo, Z., & Wu, P. (2021). ASCU-Net: Attention gate, spatial and channel attention U-Net for skin lesion segmentation. *Diagnostics*, *11*(3), 501. https://doi.org/10.3390/diagnostics11030501

[33] Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision*, 3–19. https://doi.org/10.1007/978-3-030-01234-2_1

[34] Cao, H., Bao, C., Liu, C., Chen, H., Yin, K., Liu, H., ..., & Sun, X. (2023). Attention where it matters: Rethinking visual document understanding with selective region concentration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 19460–19470. https://doi.org/10.1109/ICCV51070.2023.01788

[35] Hassan, E., Tarek, H., Hazem, M., Bahnacy, S., Shaheen, L., & Elashmwai, W. H. (2021). Medical prescription recognition using machine learning. In *Annual Computing and Communication Workshop and Conference*, 0973–0979. https://doi.org/10.1109/CCWC51732.2021.9376141

[36] Jain, T., Sharma, R., & Malhotra, R. (2021). Handwriting recognition for medical prescriptions using a CNN-Bi-LSTM model. In *International Conference for Convergence in Technology*, 1–4. https://doi.org/10.1109/I2CT51068.2021.9418153

[37] Prabu, S., & Abraham Sundar, K. J. (2023). DocPresRec: Doctor's handwritten prescription recognition using deep learning algorithm. In S. N. Kumar, S. Zafar, E. Babulak, M. A. Alam, & F. Siddiqui (Eds.), *Artificial Intelligence in Telemedicine* (pp. 33–48). CRC Press. https://doi.org/10.1201/9781003307778-4

[38] Zhou, Y., Tang, C., & Shimada, A. (2024). A novel approach: Enhancing data extraction from student handwritten notes using multi-task U-Net and GPT-4. In *International Symposium on Autonomous Systems*, 1–6. https://doi.org/10.1109/ISAS61044.2024.10552516

[39] Shende, R., Suryawanshi, P., Gajbhiye, A., Dhumane, N., & Sharma, D. (2025). Handwriting recognition & prescription scanner. *International Journal on Advanced Electrical and Computer Engineering*, *14*(1), 165–169.

[40] Aïcha Gader, T. B., & Echi, A. K. (2020). Unconstrained handwritten Arabic text-lines segmentation based on AR2U-Net. In *International Conference on Frontiers in Handwriting Recognition*, 349–354. https://doi.org/10.1109/ICFHR2020.2020.00070

[41] Karakus, O. F., Gülcü, A., & Karaca, A. C. (2025). Adapting vision transformer-based object detection model for handwritten text line segmentation task. *Journal of Innovative Science and Engineering*, *9*(1), 28–38. https://doi.org/10.38088/jise.1471047

[42] Rekha, G. R., & Siddesha, S. (2026). Categorization and content extraction in medical prescription using YOLOv8. *Emerging Electronics and Automation*, *1*(1455), 419–428. https://doi.org/10.1007/978-981-96-9554-6_33

How to Cite: Rekha G. R., Siddesha S., & Manjunath Aradhya, V. N. (2026). Treatment Regimen Segmentation from Handwritten Medical Prescriptions Using Advanced Neural Network. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA62027648