Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

Hierarchical Attention Modeling for Handwritten Prescription Recognition: A Swin Transformer-Based Solution

Swetha V Padmavathi Polisetty[1], Deepthi Godavarthi[2*]

VIT-AP University, Amaravati, Andhra Pradesh, India.

E-mail: [1]swetha.23phd7121@vitap.ac.in, [2*]deepthi.g@vitap.ac.in

Orcid ID: [1]0009-0008-2909-1336, [2*]0000-0003-0712-6899

Abstract

Many doctors' prescriptions are handwritten and often unreadable by the chemist who sells the medicine. Patients have been put in jeopardy as a result. A deep learning architecture based on the Swin Transformer for handwritten medical prescription identification has been proposed in the research paper. The proposed architecture has both a local and a global aspect. To identify character features, the local part uses handwriting features and contextual patterns through a hierarchical attention mechanism. The recommended framework is very easy to comprehend for both the handwriting style and the distortion. The BD dataset has 4,680 labeled images showing 78 different words associated with medicine. Handwritten prescriptions from doctors are included. The dataset also allows for the creation of a comprehensive preprocessing model that scales, normalizes, and applies sophisticated data augmentations that simulate real-world conditions. The test accuracy obtained was 89.0% with macro and weighted F1 scores of 0.88, performing better than some existing techniques involving CNN and CRNN. In addition, the results of the confusion matrix validate the model's ability to detect similar drug names. To demonstrate the efficiency of the model in segregating illegible handwriting from well-written handwriting, a qualitative study of the model's performance was undertaken. The researchers showed that using the transformer model, they could effectively digitize handwritten prescriptions, which can successfully curb the errors associated with the selling of medicine. The study on the semantic understanding of the obtained results through the application of optical character recognition and language models via multi-modal fusion is a future direction for research.

Keywords: Deep Learning in Healthcare, Swin Transformer, Handwritten Medical Text Recognition, Smart Prescription Digitization, Medication Error Prevention, Healthcare Automation.

1. Introduction

The increase in medical errors in the current healthcare system is largely due to incompetence among health professionals in writing prescriptions. According to research, 42% of medication errors due to poor competence levels of health professionals writing medications [1][2]. The distribution of pharmaceuticals is an important component of patient care, but poor handwriting on prescriptions is a persistent problem [3]. Pharmacists manage thousands of

* Corresponding Author

DOI: https://doi.org/10.36548/jiip.2026.1.018
Received: 10.01.2026, received in revised form: 09.02.2026, accepted: 23.02.2026, published: 05.03.2026

prescriptions daily, and this volume can cause pharmacists to have difficulty quickly interpreting and processing poorly written prescriptions, increasing the chance of error [4].

Handwriting is unique to each person and allows for a certain amount of personal expression; however, there are many reasons why a physician's handwriting may not be legible, including time constraints and high volumes of patients, which occur in a medical office or the emergency room. As a result, a physician may focus their efforts on making a diagnosis but may not take the same care in ensuring that a prescription is clear [5][6]. Due to the important role that prescriptions play in healthcare, imprecise prescriptions can lead to significant medication errors because prescriptions differ in format, and there are instances when incorrect amounts or forms of medications are written on the prescription [7].

The World Health Organization (WHO) has found a large number of prescription drug mistakes that are related to incorrect dosages due to doctors' handwriting being unreadable. On average, around 7,000 people die each year because of illegible doctors' handwriting. In 2006, the National Institute of Medical Sciences (NIMS) found that over 1.5 million Americans are harmed annually due to preventable pharmaceutical mistakes. Around 30,000 people die every year due to medical mistakes in the U.S. Additionally, a Johns Hopkins study indicated that medical mistakes are more common than stroke and respiratory diseases combined [8]. The majority of these errors are caused by a lack of usage of acronyms, improper dosage indications and illegible handwriting [9].

The Institute of Medicine reported that approximately 44,000 preventable deaths occur each year due to medical error in the United States, and 7,000 deaths can be directly attributed to illegible handwriting [10]. The study conducted at the National District Hospital in Bloemfontein aimed to determine the relative accuracy of doctors, nurses and pharmacists in understanding prescriptions. In this context, another study's results exhibited that pharmacists made the most errors when interpreting prescriptions [9][11].

Real-life instances also prove the dangerous consequences of misunderstanding drug prescriptions. A Texas study reported a case in which a cardiologist prescribed 10 mg of Plendil; however, the pharmacist misread the prescription and dispensed 20 mg instead, leading to severe complications for the patient, including death [12]. These errors highlight the challenges for both physicians and patients due to a lack of training regarding how to interpret medications and determine appropriate dosages [13].

There are two sections to a medical prescription: printed letterhead provides the physician's information such as name title, and company; handwritten information about the patient, along with the diagnosis and prescribed medications [14]. The difficulties encountered with handwritten prescriptions illustrate the need for a technological alternative to provide clearer prescriptions for reducing the occurrence of medication errors [15]. A sample prescription is shown in Figure 1 below, indicating the readability problems of handwritten medical records.

In this paper, we propose a deep learning-based recognition framework that overcomes the difficulties associated with illegible handwritten prescriptions. The proposed framework is built on the Swin Transformer network architecture, enabling quick capture of both local features of handwriting and more global, contextual features associated with the wide range of handwritten records created in the real world. We relied on a publicly available dataset, the Doctor's Handwritten Prescription BD Dataset, hosted on Kaggle, to produce, train, and evaluate the proposed model. Significant data augmentation methods were applied to enhance the model's capabilities to deal with the many variations and distortions found in handwritten prescriptions. The main goal of this effort is to improve the reliability and accuracy of prescription digitization, with the ultimate aim of increasing the safety of medications and the level of automation used in the healthcare industry.

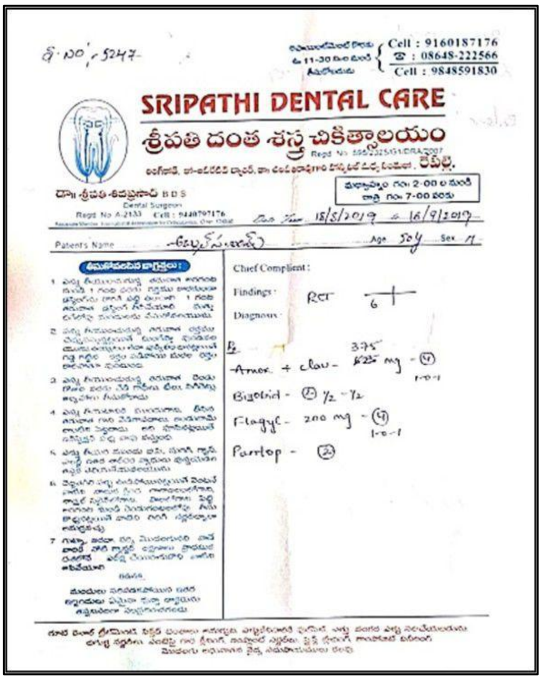

Figure 1. Sample Prescription

2. Related work

This integrated system of recognition uses machine learning to identify pharmaceutical names and dosages that have been handwritten after being prescribed. It does this by using techniques such as Optical Character Recognition (OCR) in conjunction with transformer models, along with pre-processing techniques such as eliminating noise and image size corrections to improve the overall accuracy of the identification process. There is also a mobile app that can help increase the efficiency of digitizing prescriptions for both pharmacists and patients because it is designed to be simple to use [16].

The study demonstrates how Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) models may be used to recognize and distinguish between printed and handwritten medical scripts in a variety of languages. Furthermore, these models improve the process of recognizing handwriting using Optical Character Recognition (OCR) and converting medical scripts to digital text. Finally, fuzzy search and market basket analysis are used to improve the accuracy of the transformed data, resulting in authentic and well-organized findings [5].

This study focuses on the challenge of recognizing unreadable handwriting from doctors' orders by implementing a Deep Convolutional Recurrent Neural Network (CRNN). The research found that it had a training accuracy of 76% and a validation accuracy of 72%. However, it shows that there is a need for larger datasets, different recognition methods, and better-trained neural networks in order to increase the accuracy of recognizing illegible medical prescriptions and reduce the number of medication errors that occur in healthcare [12].

This article presents a design concept for a software application using a neural network system that will assist with interpreting handwriting from doctor's prescriptions that are challenging to read because they are handwritten. The authors constructed the system using the Extended MNIST Dataset to develop a system that can improve the legibility of handwritten text and convert handwritten text into digital form. The application of this solution for digitizing prescriptions, recognizing historic documents, processing bank cheques, and interpreting handwriting in real-time was also discussed within the article [17].

A deep learning-based handwriting system using the CNN-Bi-LSTM model addresses the problem of ambiguous prescriptions. The methodology includes convolutional networks for feature extraction and bi-directional LSTMs for sequence prediction, combined with CTC (Connectionist Temporal Classification) as a decoding method for accurate character recognition. The use of curated medical corpora, along with various string-matching techniques, increased the final output recognition rate of this study. Finally, this study revealed the capacity of deep learning to boost prescription legibility, thus minimizing the likelihood of medication errors and enhancing healthcare delivery [13].

This study proposed a CRNN-based system for the accurate recognition of handwritten medical prescriptions with the aim of resolving problems related to misinterpretation that can lead to inaccuracies. It is shown that the model was developed using a combination of convolutional and recurrent layers, resulting in a model with 98% predictive accuracy. By improving legibility on prescriptions in addition to decreasing workflow and medication errors, this method ultimately improves the efficiency of the entire healthcare system [10].

In this paper, we will discuss recent improvements in deep learning by providing examples of different types of deep learning technologies used to recognize handwritten prescriptions through three main methods: Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) and Transformers. In addition, we will demonstrate how the use of YOLOv4 provides accurate localization of prescription forms in an image and implement an alternative method for sequence recognition using CRNNs with CTCs. Finally, we will highlight some existing research focused on multilingual prescription recognition, signature verification, and text-to-speech technology that improve, medication accessibility, ensure accurate medication dispensing increase efficiency in healthcare workflows, and expand access to healthcare services [4].

A Bi-LSTM based system that can recognize and interpret handwritten prescriptions made by doctors, as well as handle issues related to illegibility. The model was built using the Handwritten Medical Term Corpus, which contains 17,431 samples, and was further supplemented through the use of SRP data augmentation. The model has achieved 89.5% accuracy in prescription recognition. As a result of this significant growth in prescription legibility, the number of drug errors will be reduced and the efficiency of healthcare systems will be improved [18].

The research study developed a deep learning system to identify and classify medications using written prescriptions, aiming to improve the accuracy and speed of healthcare delivery systems. The system improves the recognition of handwritten prescriptions by applying advanced technologies such as the LayoutLMv2 and DONUT models that provide stroke enhanced recognition and use a multimodal approach. In addition, machine learning will support the optimization of the classification process with a focus on high-quality accuracy and reliability [7].

A Natural Language Processing (NLP) system is described that can be used to categorize medical prescriptions, thus helping to support the growth of healthcare digitization. This technology extracts text from scanned prescriptions and accurately classifies them using machine learning algorithms. It also interfaces with a RESTful Web Service, which helps automate this procedure. The NLP System Classifier Model was trained using a dataset that includes over 800,000 prescription records, and it performs well in categorizing diagnostic statements, enabling greater speed and accuracy in processing medical data [19].

A recommendation system developed by physicians employs Natural Language Processing (NLP) and unstructured data inputs to create a medical recommendation that increases patient safety compared to administering drugs. Estimates suggest that roughly 42% of drug errors stem from incorrect prescriptions, so this system will help physicians identify which drugs are appropriate for their patients. This system will improve diagnostic accuracy and the selection of appropriate drugs using various machine learning (ML) techniques, such as Random Forest, Bayesian Networks, and Deep Learning. By using data-driven approaches to personalize medicine, there will be fewer chances of prescription errors, thus improving the outcomes of treatment [1].

The paper describes a machine learning-based technique for identifying and digitizing handwritten medical prescriptions. This system uses Optical Character Recognition (OCR) and Named Entity Recognition (NER) to achieve a handwriting recognition accuracy rate of 64-70 %, as well as a medical data classification accuracy rate of 95-98%. This technology makes it easier to read a prescription, reduces the number of errors, and increases patient access to medical information [8].

A computerized system that classifies printed and written prescriptions has been developed using machine-learning technology. Data augmentation methods and OCR techniques were employed to separate the two categories of text accurately. The GUI will provide a mechanism for users to access pricing and other information pertaining to prescriptions, and classify the prescription text based on the writing style used. When this happens, the quality of patient care improves, as does the efficiency of the overall healthcare delivery system [14].

The development of a speech-to-text NLU-based prescribing system was incorporated into clinical practice as a means of providing an electronic method for physicians to prescribe medications by dictating them to the computer. The dictation process describes how the physician's speech is transcribed into written form and how the information provided through the transcribing process is parsed using text-to-to-speech and NLU methods to produce an electronic format that meets the e-prescribing requirements and reduces the potential for errors while increasing the legibility and accessibility of the prescription [20].

In Bangladesh, 97.1% of doctors still handwrite prescriptions. To improve the readability of these handwritten prescriptions, a machine learning (ML) technique has been proposed to enhance prescription accuracy, reduce errors in drug dispensing, and improve healthcare safety by applying automated handwriting recognition using the VGG16 model to a selected dataset [21].

This study investigates various challenges related to handwriting recognition using different applications for automatic postal system processing or other types of documents through machine learning capabilities via either offline or online forms thus enabling greater accuracy, efficiency levels, and providing additional capabilities such as allowing us to convert any non-digital information ,into a digital version [22].

A rule-based text mining system has been developed to convert free-text medical prescriptions into structured data. The proposed approach was able to identify the variability in both dosage and frequency with an accuracy of 91%. Analysis suggests that there is inherent variation in about 25% of prescriptions, which supports improved accuracy in pharmacoepidemiologic studies [23].

In this paper, we created a hybrid deep learning model to successfully identify medicinal names in handwritten prescriptions using a Mask R-CNN and Text Recognition (TrOCR) with a Multi-Head Attention Mechanism. The model performed exceptionally well with our heterogeneous dataset collected from Pakistan, yielding an impressive 1.4% character error rate, which demonstrates its capability, accuracy and real-world applicability [24].

3. Methodology

This paper proposes employing the Swin Transformer architecture to effectively recognize handwritten medical prescriptions. In order to handle the various changes that may be observed in the images of handwritten prescriptions, the suggested method has implemented adaptive training in addition to making use of significant augmentations for the dataset. Preprocessing the data, designing the model, and training the model are all steps in the proposed method's pipeline that have increased its robustness and recognition accuracy.

3.1 Data Collection

The Doctor's Handwritten Prescription BD Dataset, which includes 78 different medication names and 4,680 labelled word-level images, was used in this study. The dataset is available for the academic community to use for validation or additional research and can be downloaded from Kaggle [25]. Although the dataset is very representative of real-world prescription scenarios, recognition systems find it challenging due to the wide variety of handwritten text styles and sources of noise artifact. A variety of writing styles are represented in this dataset; some are extremely well-written, while others are noisy, hurried, distorted, and in cursive. Since the samples are all written in English and do not reflect the multilingual aspect of medical practice in many different parts of the world, there is no reason to doubt their relevance. However, it is important to note that although the dataset has an enormous variety of handwritten data, the small size of the dataset and it is based on one language alone may lead to certain biases and limitations in the results, especially in relation to medical prescribing practices all over the worldwide. The data is divided into three sets based on a stratified sampling approach, with 70%of the total number of images dedicated to the training set, 15%to the validation set, and another 15%to the testing set. The dataset, as described in the Kaggle description, is based on prescription data, with the regions of medicine names alone being cropped and marked. Therefore, consists of word-level images of medicine names alone, without including patient data and personal health information, thus anonymizing the data to the fullest.

Even though the dataset is limited to a particular linguistic environment, the handwriting features, such as cursive writing, irregular writing, and noise in the images, are common in clinical prescription writing in many parts of the world. Therefore, despite the dataset being limited to a specific linguistic environment, it can be considered appropriate for testing the effectiveness of methods for handwritten prescription recognition.

3.2 Data Preprocessing

Preprocessing of the data can improve the effectiveness of models dealing with handwritten medical prescriptions. Handwritten medical prescriptions are highly noisy and consist of many irregularities, making them a suitable candidate for data preprocessing before feeding them to any machine learning model. All the data preprocessing techniques applied during the training of the models were helpful in ensuring the accuracy of the data fed to these models.

3.2.1 Image Resizing

To maintain uniformity in all the data fed to the models, each of the prescription word images was resized to a uniform size of 224x224 pixels to ensure compatibility with the Swin Transformer model architecture.

3.2.2 Image Normalization

The ImageNet dataset's standard mean (μ = [0.485, 0.456, 0.406]) and standard deviation (σ = [0.229, 0.224, 0.225]) values were used for normalizing images. Normalizing the data yielded a consistent distribution of input data that led to more stable training and quicker convergence of the model.

3.2.3 Data Augmentation

A comprehensive data augmentation strategy was used to increase the diversity of the data used to train the model and improve its ability to generalize across a wide range of handwriting styles. It provided 5-pixel edge padding to protect the image contents during data augmentation. Random rotations of up to ±5° were used to simulate variations in handwriting styles by mimicking the variability in the angle of the handwriting. Affine transformations were applied to simulate common distortions present in handwritten text data by using small scales from 0.95 to 1.05 and translations of up to 5%. Perspective distortion was used with a probability of 30%, with a distortion scale of 0.2, to simulate cases of skewing present in handwritten data due to conditions such as the position of the document while scanning, camera-based scanning, and other minor scanning conditions. Color jittering was employed to adjust the brightness and contrast of the images by simulating varying illumination conditions. Gaussian blurring was applied with a probability of 5% to simulate low-resolution images. Random conversion to grayscale with a probability of 3% was used to simulate conditions related to low-contrast images resulting from faded ink and other scanning conditions.

3.2.4 Training Setup

In training, we used the AdamW optimizer with an initial learning rate of 3×10^{-4} and a batch size of 16. The Swin Transformer model was trained over 60 epochs, while both the CNN and CRNN reference models were trained for 40 epochs each. Additionally, a cosine annealing learning-rate schedule and label smoothing were used to help improve training stability and generalization. Finally, the model checkpoint selected for evaluation achieved the best performance on the validation data.

3.3 Proposed Methodology

The proposed framework is based on the Swin Transformer-Base (Swin-B), which is suitable for extracting both local and global contextual information from the data as well as handling the complex characteristics of handwritten prescriptions. The embedded feature tokens are generated by first preprocessing the images, dividing them into fixed-size patches, and then passing them through a linear embedding layer. The generated feature tokens are further processed through the Swin Transformer in four stages of hierarchical representations. In Stage 1, Window-based Multi-Head Self-Attention (W-MSA) is applied to extract low-level features such as strokes, edges, and basic character representations. In Stage 2, Shifted Window Multi-Head Self-Attention (SW-MSA) is applied to further enhance the spatial continuity between adjacent character areas, particularly for cursive handwriting recognition. In Stage 3, W-MSA and SW-MSA are combined, which contributes greatly to the recognition performance by learning discriminative mid-level word representations that are beneficial for enhancing the discrimination of visually similar medicine names by producing stronger attention maps for hard samples and downplaying those for easy samples. In Stage 4, the focus is on the extraction of higher-level semantic features as well as more complex spatial information. For the purpose of efficient multi-scale learning, the patch merging operation is conducted between stages, which helps increase the dimensionality of the features while reducing the spatial resolution. In the case of classification, the features are passed through the global average pooling layer followed by the fully connected softmax layer. Due to the hierarchical transformer architecture, the model can efficiently recognize different types of handwritten prescriptions for medication names. The overall Swin Transformer-based architecture for handwritten medical prescription recognition is shown in Figure 2.

3.3.1 Computational Complexity

The proposed framework employs the Swin Transformer-Base (Swin-B) backbone with an input resolution of 224 × 224 pixels. This architecture contains approximately 88 million parameters and requires about 15.4 GFLOPs per forward pass, as reported in the original Swin Transformer design. The additional classification layer introduced for recognizing 78 medicine classes adds negligible computational overhead relative to the backbone and, therefore, does not significantly affect the overall model complexity.

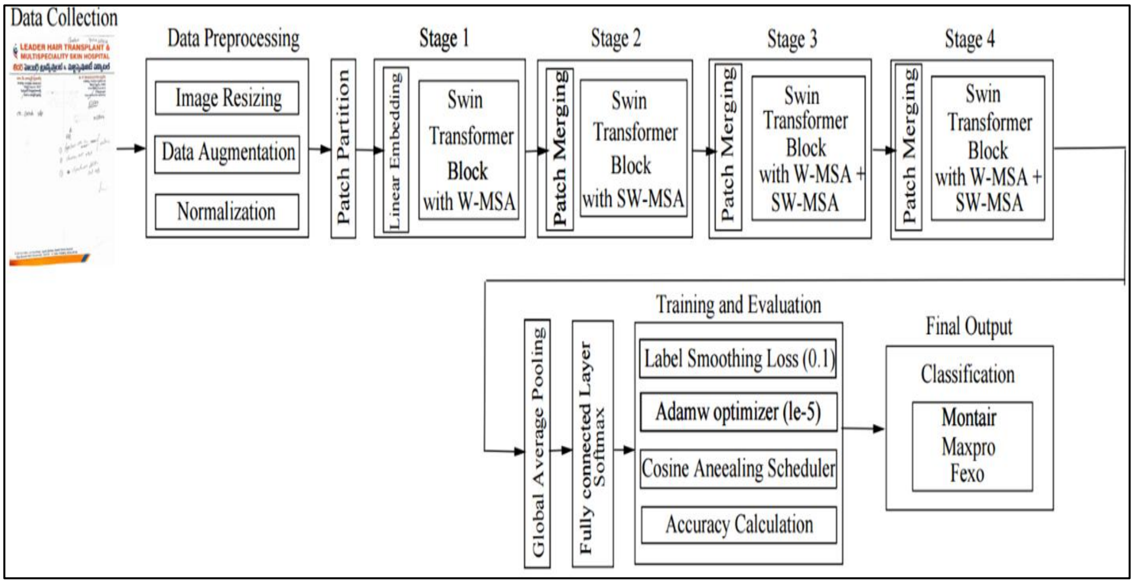

Figure 2. Handwritten Medical Prescription Recognition using Swin Transformer-Based Architecture

4. Results and Discussion

A Swin Transformer-based classification model showed a great ability to classify handwritten medical prescriptions with a test accuracy of 89.0%. Additionally, the model attained macro and weighted F1-scores of 0.88, indicating that it performed consistently and evenly well for all 78 unique medications in terms of F1 score. These results are indicative of the model's overall quality in performing classification on prescribed medications given the wide variety of handwriting variability observed in real-world applications.

4.1 Training Convergence and Performance Progression

Figure 3 shows a sequence of increasing training accuracies that result in a nearly perfect 100% training accuracy at approximately the 60th epoch. At the same time, the validation accuracy increases rapidly, peaking at around 95% at the 60th epoch. This demonstrates that the model has learned meaningful discriminative features and did not become overfitted to either set of data. The downward trend exhibited by the training and validation losses shown in Figure 4 also supports this conclusion. Overall, the combination of the AdamW optimizer, cosine annealing learning rate scheduler, and label smoothing regularization provided a strong optimization process, leading to greater stability in convergence and higher generalization capabilities.

The early saturation of training accuracy while validation accuracy continues to improve can be attributed to the combined effect of very strong regularization and the extensive use of data augmentation during training. The use of label smoothing, very aggressive geometric and photometric augmentations, and the AdamW optimizer with a cosine annealing learning rate scheduler restrict the model's potential to memorize training samples, causing early saturation of training accuracy. Conversely, validation accuracy continues to improve as the model learns progressively more robust, discriminable feature representations that generalize well to previously unseen data. This type of training behavior indicates the presence of effective generalization, rather than overfitting or underfitting, and lends stability to the proposed Swin Transformer-based methodology.

Based on the observations during the training process, it was decided to stop training at 60 epochs because at that point training loss and validation loss stopped improving. After 60 epochs, there had been no noticeable increase in validation performance; thus, any additional training would not provide valid data for future performance increases and could lead to excessive costs in computing resources or increase the chance of overfitting.

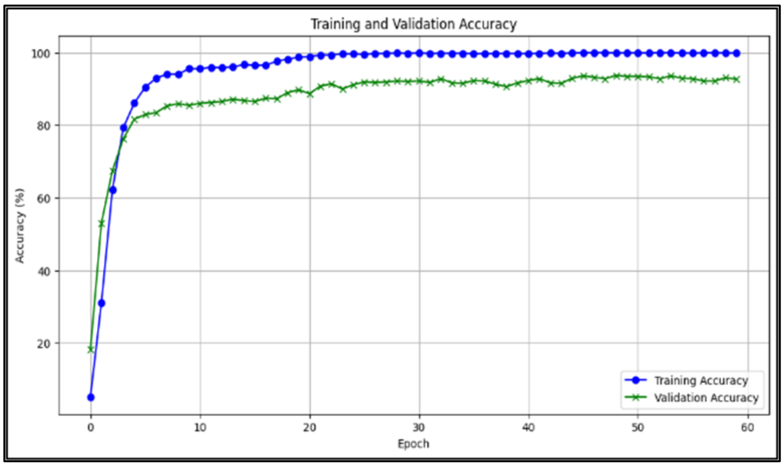

Figure 3. Training and Validation Accuracy Over Epochs

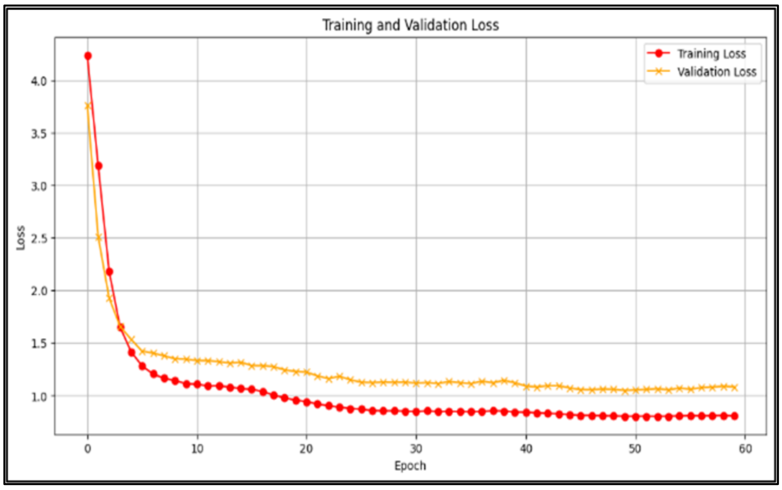

Figure 4. Training and Validation Loss Over Epochs

4.2 Misclassification Trends and Confusion Matrix Evaluation

Figure 5 depicts a confusion matrix illustrating the 10 most misclassified medicine classes to further explain the model limitations presented. The model misclassified several different medicines, such as Esoral, Renova, and Rivotril, with high frequencies. The model likely misclassified these medicines due to similarities in their physical characteristics, ambiguous character strokes, and handwriting artifacts that resulted in difficulty disambiguating these categories, especially when written in low-quality or very cursive handwriting.

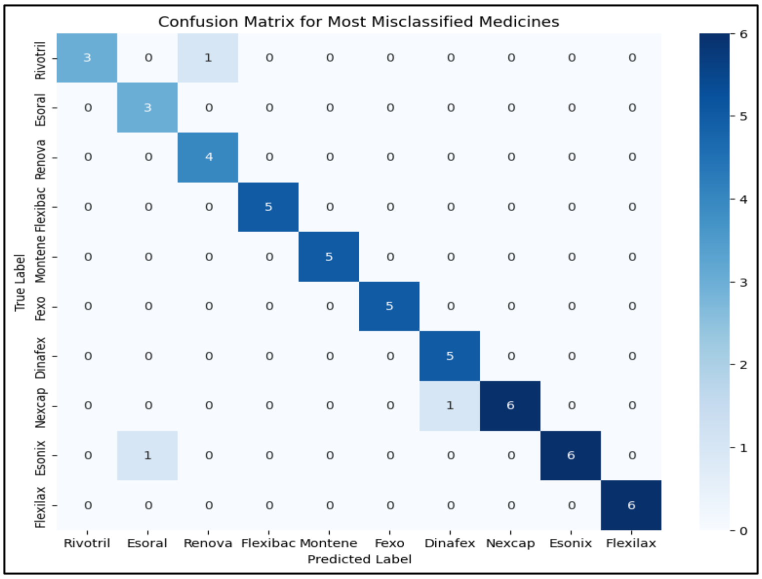

Figure 5. Confusion Matrix of the Top 10 Most Misclassified Medicine Names

4.3 Comparative Analysis with CNN and CRNN Baselines

To better contextualize the performance of the proposed Swin Transformer model, we conducted a direct comparison with two widely used baselines: a Convolutional Neural Network (CNN) and a Convolutional Recurrent Neural Network (CRNN). All models were trained and tested using the same preprocessing methods and stratified splitting (the dataset was composed of 3,276 training images, 702 validation images, and 702 test images with 78 categories) of the Doctor's Handwritten Prescription BD Dataset. The Swin Transformer had the highest validation accuracy (95%), test accuracy (89%). The CNN had a validation accuracy of 91.92% and a test accuracy of 80.51%. The CRNN baseline had the lowest validation accuracy of 80.51% and a test accuracy of 60.90%. Therefore, the Swin Transformer is superior to both the CNN and CRNN when modeling complex handwritten characters, but the CNN and CRNN models have difficulty with cursive writing, overlapping characters, and distortion caused by noise. The Swin Transformer demonstrates the benefit of using hierarchical attention mechanisms for accurate automated handwritten prescription recognition, as shown in Table 1.

Table 1. Performance Comparison of Swin Transformer, CNN, and CRNN Baselines

Model	Train Accuracy (%)	Peak Validation Accuracy (%)	Test Accuracy (%)	Key Observation
CRNN	96.76	80.51	60.90	Performed poorly on noisy or cursive scripts; sensitive to handwriting variability.
CNN	99.58	91.92	80.51	Outperformed CRNN but less robust to handwriting distortions and overlapping strokes.
Swin Transformer (Proposed)	100.00	95.00	89.0	Best overall performance, effectively captured complex handwriting patterns.

Table 1 presents evidence supporting the claim that the Swin Transformer consistently outmatches the baseline models when evaluated on the same dataset and under the same conditions. Although the CNN delivered adequate performance, it was less consistent than the Swin Transformer on handwritten document images with distortions. The CRNN performed poorly compared to all of the other models and had a particularly difficult time processing cursive writing and noisy images. This illustrates that the Swin Transformer not only provides better performance than the other models, but it is also superior at recognizing complex features of handwriting.

4.4 Per-Class Performance Metrics Evaluation

Figures 6(a), 6(b) and 6(c) provide an analysis of the performance of the Swin Transformer across classes. Each figure shows class-wise performance metrics for the 10 most commonly misclassified medicines in the dataset. The model exhibited exceptional performance for Flexibac, Nexcap, and Esonix, with these three classes scoring well on all three-evaluation metrics, indicating the model's ability to identify these medicines despite the variability in handwriting. Conversely, Esoral and Rivotril had a lower-than-expected recall value, suggesting that the model is less likely to consistently distinguish between these two medicines when handwritten forms exhibit variability. Reasons for this model performance include potential character ambiguity, similar stroke patterns, and image artifacts. The results demonstrate the overall effectiveness of the model in recognizing medicines; however, certain areas of performance represent opportunities for improvement in the model's ability to recognize complex handwritten input as typically encountered on actual prescriptions.

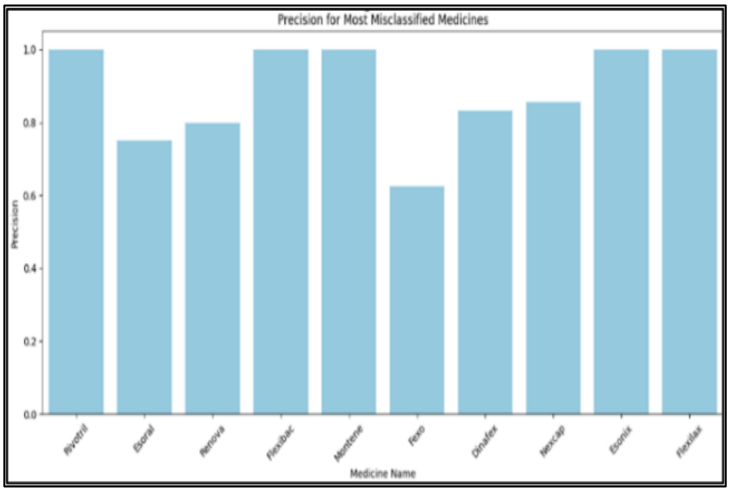

Figure 6(a). Precision Scores for the Most Misclassified Medicines

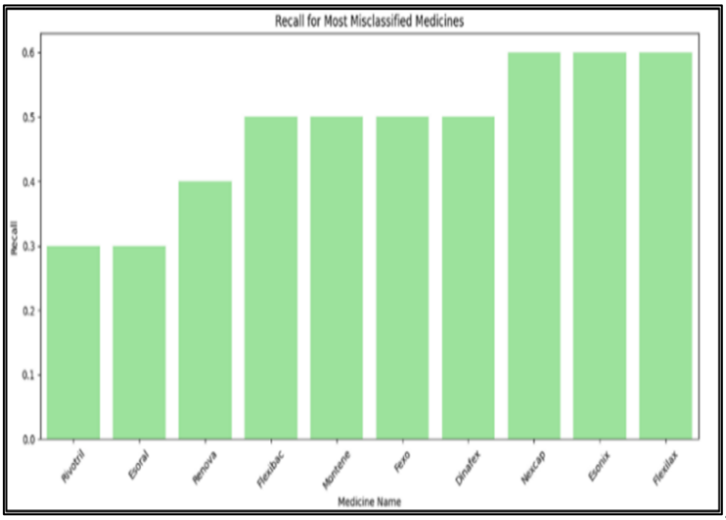

Figure 6(b). Recall Scores for the Most Misclassified Medicines

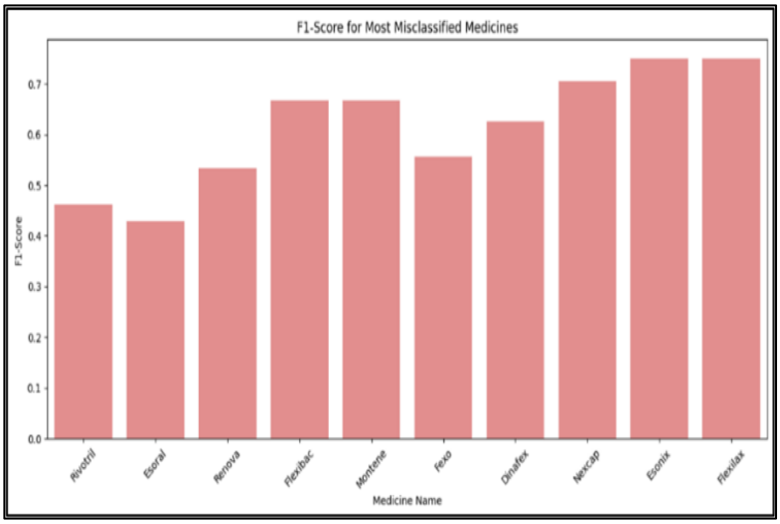

Figure 6(c). F1-Scores for the Most Misclassified Medicines

4.5 Analysis of Correct and Incorrect Predictions

To continue examining the Swin Transformer classification performance, Figure 7 provides an overview of how correctly and incorrectly predicted samples are distributed among those classes that resulted in the most misclassifications. The data indicates that although the model performs well for the majority of classes, there are some classes that have a relatively high number of errors when making predictions about samples. Therefore, while the results show good overall generalization of the Swin Transformer model, they suggest that further improvements will need to be made in order to distinguish between similar or ambiguous drug names. To provide additional context to the numbers presented in Figure 7, Figure 8 illustrates examples of both correctly classified (green) and incorrectly classified (red) drugs. These examples illustrate successful cases, such as clear or somewhat slanted handwriting, as well as unsuccessful cases, including very cursive handwritten stroke patterns, overlapping letters, or other noise-related artifacts that create ambiguity between visually similar drug names like Esoral and Rivotril.

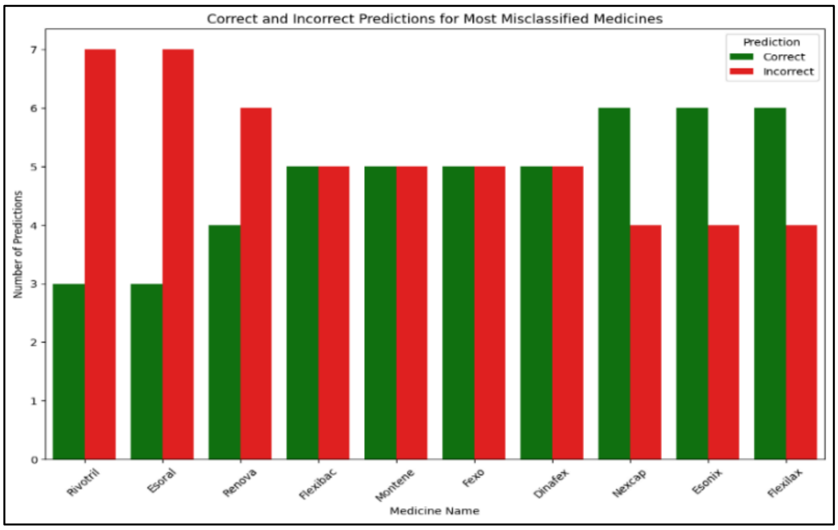

Figure 7. Correct vs Incorrect Predictions on Misclassified Classes

For the most commonly misidentified medicine names, predictions were counted as correct (green) or incorrect (red). The results show the degree to which handwriting variability, such as cursive writing, overlapping lines, or noise artifacts, has affected the prediction performance of the model. The counts of how this occurred are represented in visual samples shown in Figure 8.

4.6 Visual Inspection of Predictions

The Swin Transformer's predictions can be inspected qualitatively through the images shown in Figure 8. Correctly predicted prescriptions are shown in green, and incorrectly predicted prescriptions are denoted in red. The visual analysis indicates that the Swin Transformer produced many successful predictions when given images with legible handwriting and well-organized prescriptions. Conversely, when samples exhibited significant distortion, poor handwriting, or overlapping strokes, there were many incorrectly predicted samples. Analysis of the qualitative data also demonstrates a high level of challenge with respect to collecting features in an effective and reliable manner from various types of handwritten medical prescriptions. These highly successful and unsuccessful visual examples allow for a clearer relationship between successful and unsuccessful qualitative feature extraction and the quantitative measures that resulted in both: namely, the characteristics that exist in handwritten prescriptions.

Figure 8. Qualitative Visualization of Correct and Incorrect Predictions

Word images are represented by prescriptions showing the correct classifications (green highlights) and misclassifications (red highlights). Numerous examples demonstrate the ability of handwritten text to be successfully classified when written legibly and in clear writing, while other examples illustrate how poorly written, highly cursive, distorted, or noisy handwriting can create visual confusion among similar drug names, such as Renova, Esoral, and Rivotril; thus complementing the quantitative results provided in Figure 7 by showing the strengths and weaknesses of the model.

4.7 Additional Evaluation Metrics and calibration

To complement overall accuracy and macro/weighted F1 scores, we report additional agreement and calibration metrics, along with probability-quality diagnostics. The full confusion matrix in Figure 9 shows that most errors are concentrated within a limited number of visually similar medicine classes. The reliability diagram in Figure 10 (ECE ≈ 0.067) indicates that the Swin Transformer’s confidence scores are generally well-calibrated. An

Expected Calibration Error (ECE) of 0.067 reflects a small average gap between the model's predicted confidence and its observed accuracy. Consequently, calibration can be considered strong, as this value indicates well-aligned confidence estimates rather than overconfident predictions, which is acceptable for a multi-class handwritten prescription recognition task used in a decision-support setting. In medical decision-support systems, this level of calibration is regarded as appropriate, particularly for non-autonomous applications where model outputs are reviewed by healthcare professionals rather than acted upon automatically. Such calibration reduces the risk of overconfident, incorrect predictions and supports safer downstream verification. When combined with high top-k accuracy, robust ROC-AUC values, and human-in-the-loop validation, the reported ECE value supports informed clinical decision assistance rather than direct medical decision-making. The confidence histogram presented in Figure 11 shows that the majority of predictions are made with high confidence, while the macro One-vs-Rest ROC curve in Figure 12 demonstrates strong class separability, with an AUC of approximately 0.992.

Balanced accuracy, Cohen's kappa (k), Matthews correlation coefficient (MCC), top-3 and top-5 accuracies, log loss, and Brier score are presented in Table 2, providing a more comprehensive evaluation of the model's performance beyond overall accuracy.

Table 2. Additional Evaluation Metrics for the Swin Transformer

Metric	Value
Accuracy	0.890
Balanced Accuracy	0.890
Top-3 Accuracy	0.944
Top-5 Accuracy	0.955
Macro Precision	0.907
Macro Recall	0.890
Macro F1-score	0.886
Weighted Precision	0.907
Weighted Recall	0.890
Weighted F1-Score	0.886
Cohen's kappa (κ)	0.888
Matthew's correlation coefficient (MCC)	0.889
ROC AUC (Macro OvR)	0.992
ROC AUC (Weighted OvR)	0.992
Log Loss	0.681
Brier Score	0.175
Expected Calibration Error (ECE)	0.067

The overall recognition error of the proposed framework appears to be about 11%, based on a reported test accuracy of 89.0%. Errors due to confusion arise chiefly from visually similar names of medications as well as handwriting that is substantially cursive or distorted. Misclassification is systematic rather than random, indicative of some level of ambiguity regarding handwritten prescriptions.

Top-k accuracy metrics were introduced to account for the ambiguity of handwritten prescriptions, where many medicine names that look similar or different handwriting styles can generate multiple possible outputs. Top-3 and Top-5 were chosen to provide a more general reflection of what happens in clinical and pharmacy settings, where a pharmacist or automated decision-support system can use the top-3 or top-5 probable candidates as a basis to make a decision rather than relying solely on the top-1. Top-3 accuracy accounts for situations where the correct medicine is among the top-3 probable alternatives, and Top-5 provides an additional "margin of error" for medicines verified by error-tolerant methods. Both Top-3 and Top-5 are widely used in multi-class recognition problems and provide a reasonable compromise of usability, safety, and diagnostic confidence.

This results table includes information on additional performance metrics beyond accuracy, such as: balanced accuracy, Cohen's kappa (k), Matthews correlation coefficient, top-k accuracy of each model, Brier Score, log loss, and ROC-AUC. This information shows that our model has strong agreement with other models, is well calibrated and provides decisions with high confidence based on its classification accuracy.

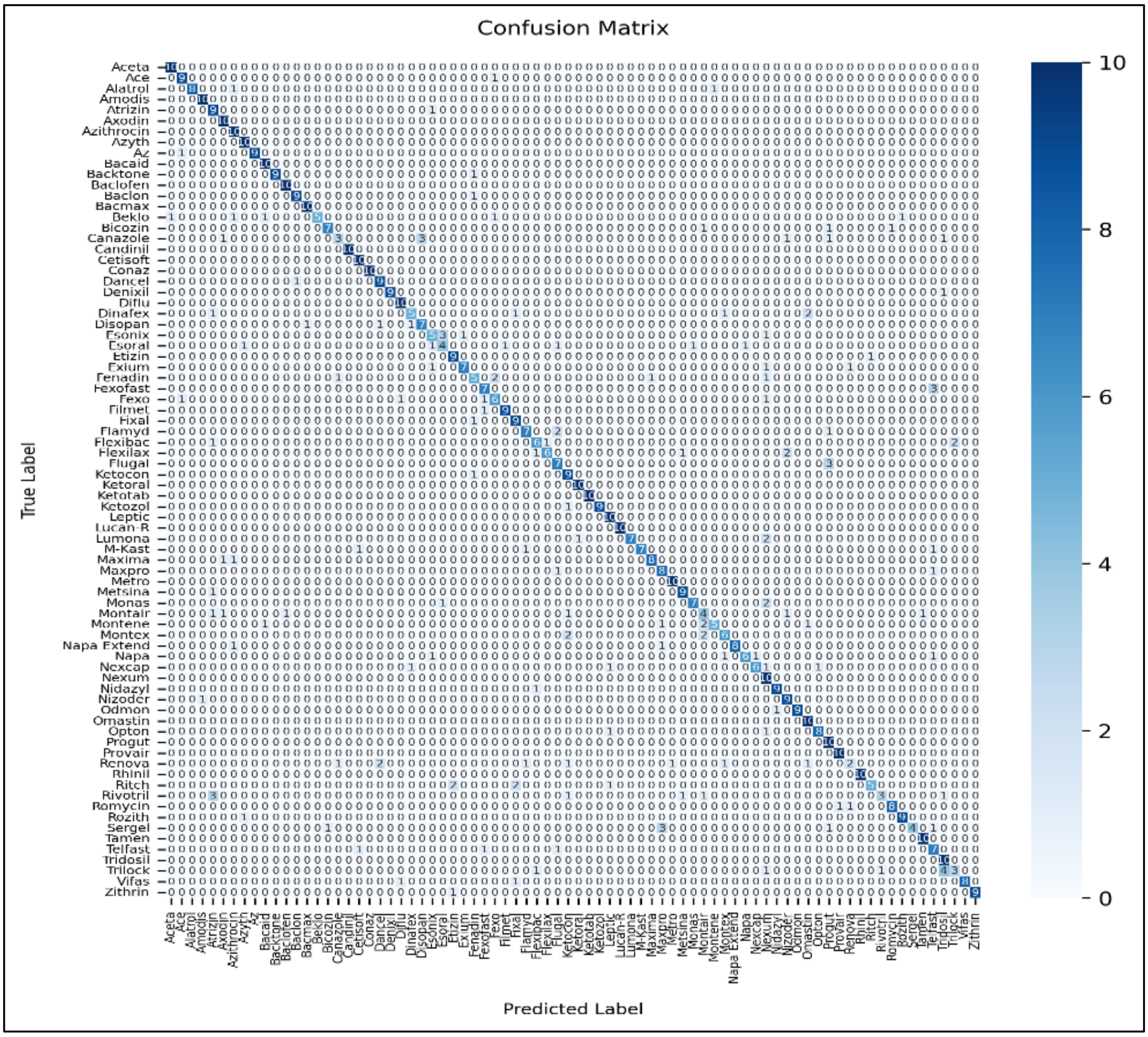

Figure 9. Full Confusion Matrix for the Swin Transformer Model

In the comprehensive confusion matrix for all 78 classes of medicines, it is clear that most of the predictions are found along the overall diagonal. This indicates very good performance in terms of classifying these medicines as well as clear evidence of separation among these different classes of medicines. The misclassifications do not follow a totally random distribution but seem to be primarily focused within rather limited collections of medicines with visually similar names. Those misclassifications are more likely to occur among classes whose names contain similar characters or character configurations, stroke characteristics, or word lengths, especially when written in a cursive, distorted, or very poor-quality manner. Additionally, while there are some classes with nearly perfect recognition and very little confusion, the sparsely populated off-diagonal cells within the confusion matrix imply that the remainder of errors in recognition were caused by visual ambiguities that exist in handwritten prescriptions rather than true overlap between the respective medicine classes or instability of the model. These errors are primarily caused by overlapping strokes for a given character, irregular continuity of strokes for a given character, and variability in the method by which handwriting is produced, ultimately decreasing the visible distinctiveness between specific names of medicines.

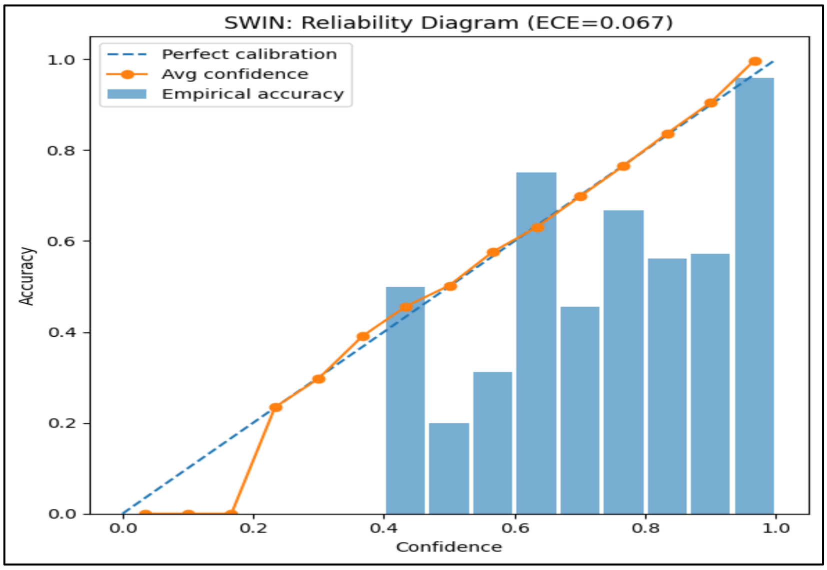

Figure 10. Reliability Diagram with Expected Calibration Error (ECE ≈ 0.067)

A reliability diagram is used to compare predicted confidence to empirical accuracy. The ECE value is low demonstrates that the predicted probabilities are well-calibrated, meaning the confidence score is closely correlated with the observed accuracy.

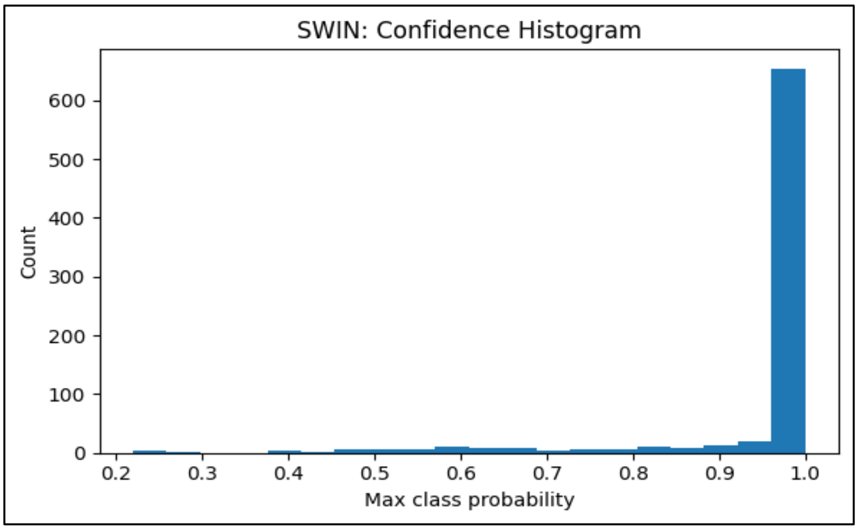

Figure 11. Confidence Histogram for Swin Transformer Predictions

Class probability distributions across the entire dataset, most of which show a high level of confidence, indicate that the model was able to generate confident and conclusive outputs from all writing styles and samples.

High-confidence misclassification in a clinical environment is minimized via human verification whereby model output assists clinical decision-support and does not provide autonomous decisions. When assigned high confidence by the system, the assistant outputs the high-confidence prediction with corresponding alternative top-k candidates along with an associated confidence level. This allows pharmacists to compare the high-confidence output from the model against their domain knowledge and prescription use case. If there is any ambiguity or uncertainty, the final authority for verification is manual, meaning that high-confidence predictions are never automatically accepted as correct. This workflow means that even high-confidence errors will not precipitate a clinical action, maintaining patient safety while utilizing the efficiency benefits of automated prescription recognition.

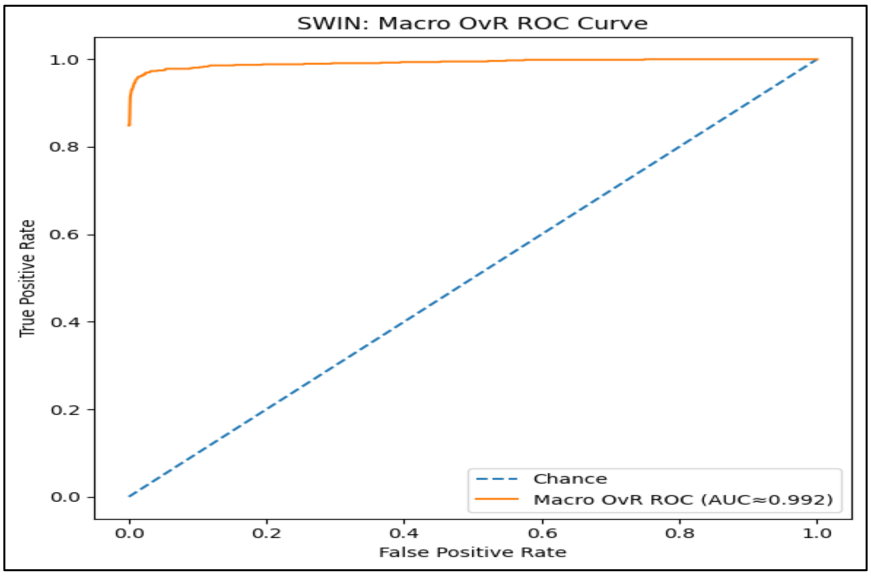

Figure 12. Macro One-vs-Rest ROC Curve (AUC ≈ 0.992)

The mean ROC curves across all class types display a clear boundary between each of the different class-type combinations. The high AUC value demonstrates that the Swin Transformer can reliably differentiate between correct and incorrect classes at many different threshold settings.

4.8 Generalization Considerations

This study primarily addresses domain shifts using data augmentation and robust feature learning. All training pipeline augmentations consist of geometric and photometric transforms simulating real-world variances present in common clinical prescriptions, such as stroke warping, slant orientation differences, noise, and lighting discrepancies. The Swin Transformer's hierarchical attention mechanism also helps the model capture both fine details of handwritten text and overall structural patterns, which increases overall robustness against unknown variations from unseen domains. Even though these methods provide increased resistance to domain shifts in practice, additional robust validation with diverse datasets will need to be conducted in order to fully assess cross-domain generalization.

The clinical importance of this study is that the correct names of medicines are emphasized, as they constitute a key step in prescription interpretation and drug safety. Identification of the prescribed medication is a prerequisite for checking dose, frequency, and combination of medications in clinical and pharmacy processes. It is characterized as a decision-support system and should help medical staff avoid mistakes that originate from illegible handwriting, not replace clinical judgement. Dosing schedules and drug interactions would generally be corroborated by pharmacist review or e-records and could potentially be included in the system with further extensions.

The Swin Transformer performed well on the Doctor's Handwritten Prescription BD Dataset, but there has been no evaluation yet regarding its applicability to other geographic areas or languages. There are differences between how different authors write and how various types of abbreviations and medical terminologies are used at the clinical facility level, which could create unexpected results that would impact real-life usage. The next stage of research will be to conduct additional evaluations of multilingual datasets from various healthcare systems, using transfer learning and domain adaptation techniques to improve the original performance.

The proposed framework is modular, transformer-based, and scalable for large deployments and clinical use. Patch-wise processing allows for efficient batch inference of large batches of patient prescriptions. The Swin Transformer backbone enables rapid batch processing on standard clinical server systems or cloud-based platforms. Therefore, the system will support high-throughput processing of prescription images. The framework can also incrementally incorporate new classes of medicine without changing the general framework and can be deployed in phases across many different healthcare settings.

5. Conclusion

The Swin-Transformer-Prescription model is a machine learning-based technique proposed to solve the handwritten prescription drug recognition problem using the hierarchical attention mechanism. The implementation of crucial pre-processing steps, such as image normalization and image augmentation, for the proposed deep learning-based model improved its performance, achieving an accuracy of 89.0%, a macro F1-score of 0.88, and a weighted F1-score of 0.88 across a total of 78 classes of medications. The training process, which included label smoothing, the AdamW optimizer, and a cosine annealing-based learning rate scheduler, was effective. However, the model faced difficulties with words that were too similar or poorly written. To enhance the model's performance, the dataset may be expanded to include multiple modalities. Additionally, the dataset may be extended to accommodate prescription scripts written in different languages. It should also be tested on various types of handwriting to evaluate how well the model performs.

Data Availability

We utilize the publicly available Doctor's Handwritten Prescription BD Dataset (Mamun et al., n.d., Kaggle). This dataset was neither created nor owned by the authors of this paper. Only publicly available images with no patient identifiers have been used. Each experiment and base comparison used an equal split of 70% for training, 15% for validation, and 15% for testing. The entire dataset can be downloaded directly from at Kaggle (Doctor's Handwritten Prescription BD Dataset, n.d.).

References

[1] Kumar, C. Srinivasa, Ranga Swamy Sirisati, Viswanath Gudditti, Kanusu Srinivasa Rao, and Ratna Kumari Challa. "A Smart Recommendation System for Medicine Using Intelligent NLP Techniques." In 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), IEEE, 2022, 1081-1084.

[2] World health organization, “Medication Without Harm.” [Online]. Available: https://www.who.int/initiatives/medication-without-harm

[3] Modi, Tasneem, Ntandoyenkosi Khumalo, Rubina Shaikh, Zelna Booth, Stephanie Leigh-de Rapper, and Gillian Dumsile Mahumane. "Impact of Illegible Prescriptions on Dispensing Practice: A Pilot Study of South African Pharmacy Personnel." Pharmacy 10, no. 5 (2022): 132.

[4] Edlabadkar, Sania Ravindra, and Pranjali P. Joshi. "Accurate Medicine Recognition and Audio Conversion for Handwritten Prescriptions." In 2024 International Conference on Emerging Smart Computing and Informatics (ESCI), IEEE, 2024, 01-06.

[5] Pavithiran, G., Sharan Padmanabhan, Nuvvuru Divya, V. Aswathy, and B. Chandar. "Doctor's Handwritten Prescription Recognition System in Multi-Language Using Deep Learning." In 2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), IEEE, 2022, 1-5.

[6] Abdullahi, Abba Khalid, Ibrahim Senire Fatima, Umar Abdurrahaman, Salihu Isa Sa’adatu, Abubakar Bukhari Hafsat, Haruna Abdullahi Abdulrasheed, Umar Abubakar et al. "Assessment of Legibility of Handwritten Prescriptions and Adherence to WHO Prescription Writing Guidelines in Ahmadu Bello University Teaching Hospital Zaria–Kaduna State, Nigeria." Innovations in Pharmacy 14, no. 1 (2023): 10-24926.

[7] Zia, Javeria, Usman Habib, and Muhammad Asif Naeem. "Extraction and Classification of Medicines from Handwritten Medical Prescriptions." In 2023 18th International Conference on Emerging Technologies (ICET), IEEE, 2023, 104-109.

[8] Kulathunga, Dinuka, Chamika Muthukumarana, Umindu Pasan, Chamudika Hemachandra, Muditha Tissera, and Hansi De Silva. "Patientcare: Patient Assistive Tool with Automatic Hand-Written Prescription Reader." In 2020 2nd International Conference on Advancements in Computing (ICAC), vol. 1, IEEE, 2020, 275-280.

[9] Wijewardena, W. R. A. D. "Medical Prescription Identification Solution." PhD diss., 2021.

[10] Achkar, Roger, Khodor Ghayad, Rayan Haidar, Sawsan Saleh, and Rana Al Hajj. "Medical Handwritten Prescription Recognition Using CRNN." In 2019 International Conference on Computer, Information and Telecommunication Systems (CITS), IEEE, 2019, 1-5.

[11] Brits, Hanneke, A. Botha, L. Niksch, K. Venter, R. Terblanché, and G. Joubert. "Illegible Handwriting and Other Prescription Errors on Prescriptions at National District Hospital, Bloemfontein." Professional Nursing Today 21, no. 2 (2017): 53-56.

[12] Fajardo, Lovely Joy, Niño Joshua Sorillo, Jaycel Garlit, Cia Dennise Tomines, Mideth B. Abisado, Joseph Marvin R. Imperial, Ramon L. Rodriguez, and Bernie S. Fabito. "Doctor’s Cursive Handwriting Recognition System Using Deep Learning." In 2019 IEEE 11th international conference on humanoid, nanotechnology, information technology, communication and control, environment, and management (HNICEM), IEEE, 2019, 1-6.

[13] Jain, Tavish, Rohan Sharma, and Ruchika Malhotra. "Handwriting Recognition for Medical Prescriptions Using a CNN-Bi-LSTM Model." In 2021 6th International conference for convergence in technology (I2CT), IEEE, 2021, 1-4.

[14] Dhar, Dibyajyoti, Avishek Garain, Pawan Kumar Singh, and Ram Sarkar. "HP_DocPres: A Method for Classifying Printed and Handwritten Texts in Doctor's Prescription." Multimedia Tools and Applications 80, no. 7 (2021): 9779-9812.

[15] Richard Molitor, "Tracking e-Script Errors in a Community Pharmacy," Pharm. Technol., 2025.

[16] Hassan, Esraa, Habiba Tarek, Mai Hazem, Shaza Bahnacy, Lobna Shaheen, and Walaa H. Elashmwai. "Medical Prescription Recognition Using Machine Learning." In 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), IEEE, 2021, 0973-0979.

[17] Shaw, Utkarsh, Rishab Mamgai, and Isha Malhotra. "Medical Handwritten Prescription Recognition and Information Retrieval Using Neural Network." In 2021 6th International conference on signal processing, computing and control (ISPCC), IEEE, 2021, 46-50.

[18] Tabassum, Shaira, Ryo Takahashi, Md Mahmudur Rahman, Yosuke Imamura, Luo Sixian, Md Moshiur Rahman, and Ashir Ahmed. "Recognition of Doctors' Cursive Handwritten Medical Words by Using Bidirectional LSTM and SRP Data Augmentation." In 2021 IEEE technology & engineering management conference-Europe (TEMSCON-EUR), IEEE, 2021, 1-6.

[19] Carchiolo, Viincenza, Alessandro Longheu, Giuseppa Reitano, and Luca Zagarella. "Medical Prescription Classification: A NLP-based Approach." In 2019 Federated Conference on Computer Science and Information Systems (FedCSIS), IEEE, 2019, 605-609.

[20] Shaikh, Danish Ali, Burhanuddin Fatehi, Atif Khan, Amaan Shaikh, and Nafisa Mapari. "Voice Prescription Using Natural Language Understanding." In 2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), IEEE, 2022, 1-6.

[21] Mia, Abdur Rahim, Mohammad Abdullah-Al-Sajid Chowdhury, Abdullah Al Mamun, Aurunave Mollik Ruddra, and Nawshin Tabassum Tanny. "A Deep Neural Network Approach with Pioneering Local Dataset to Recognize Doctor's Handwritten Prescription in Bangladesh." In 2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS), IEEE, 2024, 1-6.

[22] Vashist, Prem Chand, Anmol Pandey, and Ashish Tripathi. "A Comparative Study of Handwriting Recognition Techniques." In 2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM), IEEE, 2020, 456-461.

[23] Karystianis, George, Therese Sheppard, William G. Dixon, and Goran Nenadic. "Modelling and Extraction of Variability in Free-Text Medication Prescriptions from an Anonymised Primary Care Electronic Medical Record Research Database." BMC medical informatics and decision making 16, no. 1 (2015): 18.

[24] Ali, Usman, Sahil Ranmbail, Muhammad Nadeem, Hamid Ishfaq, Muhammad Umer Ramzan, and Waqas Ali. "Leveraging Deep Learning with Multi-Head Attention for Accurate Extraction of Medicine from Handwritten Prescriptions." In 2024 International Conference on Robotics and Automation in Industry (ICRAI), IEEE, 2024, 1-6.

[25] Doctor's Handwritten Prescription BD dataset, "Doctor's Handwritten Prescription BD dataset." [Online]. Available: https://www.kaggle.com/datasets/mamun1113/doctors-handwritten-prescription-bd-dataset