International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. , No. (Mon-20..)

http://dx.doi.org/10.12785/ijcds/XXXXXX

A Deep Learning Model for Medical Prescription Transcription: Integrating CNN and OCR for Enhanced Accuracy

Maheen Islam[1], Mahamudul Hasan[1] and Nishat Vasker[1]

[1]*Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh*

Received Mon. 20, Revised Mon. 20, Accepted Mon. 20, Published Mon. 20

Abstract:
Inaccurate interpretation of handwritten medical prescriptions is a pressing problem in healthcare, often leading to medication errors and adversely affecting patient safety. The complexity of deciphering diverse handwriting styles necessitates an automated, accurate transcription solution. This research addresses the critical question: How can deep learning improve the accuracy and efficiency of medical prescription transcription? We developed an advanced deep learning model combining Convolutional Neural Networks (CNN) and Optical Character Recognition (OCR) to accurately transcribe handwritten prescriptions. The methodology involved curating a dataset of prescription images, preprocessing for optimal deep learning application, and training the model to recognize and transcribe text. Our model achieved a significant breakthrough with an accuracy of 91.80% for English and 87.50% for Bangla scripts, demonstrating a robust ability to handle real-world prescription variability. The results affirm the effectiveness of integrating CNN and OCR in solving the problem of prescription transcription. The goals of enhanced patient safety and streamlined healthcare documentation have been substantially achieved. The practical implementation of this model has the potential to drastically reduce medication errors, contribute to theoretical advances in AI applications in healthcare, and bear significant ethical implications by improving patient outcomes. This research presents a novel approach to prescription transcription, offering a valuable tool for healthcare professionals. It sets a new precedent in medical documentation, paving the way for future innovations and serving as a benchmark for similar applications in healthcare technology.
Keywords: Deep Learning, Medical Prescription Transcription, Convolutional Neural Networks, Optical Character Recognition, Medication Error Prevention, Healthcare Documentation, Patient Safety, CNN, OCR

1. INTRODUCTION

In the field of healthcare, accurately deciphering doctors' handwritten prescriptions is of utmost importance, as misinterpretation can result in medication errors that can lead to serious harm to patients. The challenge is compounded by the variety and complexity of handwriting styles, which often make it difficult to read the written instructions. To improve the accuracy and efficiency of prescription transcription, this research uses deep learning by employing Convolutional Neural Networks (CNN) and Optical Character Recognition (OCR) to bridge the gap between analog prescriptions and digital accuracy.

Our investigation revolves around several research questions that guide the development of our system. Our primary question is: How can deep learning technologies be used to create a system that transcribes handwritten medical prescriptions accurately, thereby enhancing patient safety and simplifying the prescription management process? Secondary questions focus on the system's ability to adapt to diverse handwriting styles, the efficiency of the transcription process, and the accuracy of text recognition.

The aim of this research project is to develop an intelligent system that can accurately transcribe handwritten medical prescriptions, irrespective of variations in handwriting. We intend to create a dataset of prescription images and employ advanced image preprocessing techniques to train a model that can not only recognize text with precision but can also integrate smoothly with existing healthcare documentation workflows.

We have accomplished several milestones in the field of healthcare. For instance, we have developed a deep learning model that outperforms the existing benchmarks in identifying both English and Bangla prescriptions. Our extensive dataset is an accurate representation of real-world scenarios. Moreover, we have designed a user-friendly application interface that simplifies the transcription process for healthcare professionals. We have also evaluated the model's performance, highlighting its practical benefits and the potential for reducing medication errors.

This paper is structured to guide the reader through our research and development process in a logical progression. After the introduction, we will discuss related works that have laid the foundation in the field of prescription recognition. We will then explain our methodology, which includes dataset preparation and model training. Later sections will cover the results and their impact on healthcare documentation, followed by a discussion of the system's project management features and a presentation of the user interface. Finally, we will reflect on our findings, contributions to the field, and future research directions.

2. Related Works

Deep learning's influence has permeated a variety of disciplines, particularly in healthcare([1],[2]), where its applications are both crucial and sensitive. [3] critically assessed deep learning's "black box" nature, highlighting the difficulties in demystifying the internal mechanics of machine-based decisions—a pressing issue in the context of medical ethics and accountability. Simultaneously,[4] broke new ground by devising an advanced deep learning([5]) model adept at predicting drug-drug interactions (DDIs), surpassing the capabilities of existing algorithms.

As the call for transparency in AI grows louder,[6] stepped forward with a pioneering approach, integrating a type-2 fuzzy logic system into a deep learning framework to enhance model interpretability without sacrificing significant predictive power. This initiative toward explainable AI dovetails with [7] work on medical image segmentation and [8] exploration into prognostic models for lung cancer survival—both underscore the versatility and depth of deep learning applications in medical diagnostics and treatment planning.

The scope of deep learning extends beyond diagnosis and interaction prediction([9]), as illustrated by [10], who examined its potential to automate the summarization of vast health information, catering especially to consumer-centric health inquiries. In the arena of radiology,[11] emphasized the burgeoning field of multimodal deep learning, underscoring the synergistic potential of combining datasets from disparate medical specialties to refine patient assessments.

The utility of deep learning has also been explored in affective computing, with Akbulut [12] presenting a hybrid affective model that discerns human emotions with a notable degree of accuracy, further cementing deep learning's role in nuanced, non-diagnostic contexts. Concurrently,[13] surveyed the landscape of fatigue detection, spotlighting machine learning's advancements in identifying a critical physiological state with wide-ranging health implications.

Venturing into dental medicine, [14] compiled a primer on deep learning tailored for dental professionals, encapsulating various methodologies and their tangible outcomes. This body of work collectively demonstrates deep learning's robust capabilities in enhancing medical practices and patient care. Nonetheless, the quest for model interpretability persists as a central challenge—an endeavor that continues to drive innovation at the intersection of medical practice and artificial intelligence.

The integration of Artificial Intelligence (AI) into the healthcare sector([15]) represents a watershed in the field's evolution, epitomizing a shift towards more data-driven and patient-centric approaches. The proliferation of AI applications is varied and impactful, ranging from the nuanced crafting of personalized medical prescriptions to the critical surveillance against fraudulent activities within healthcare systems.

[16] delve into the complexities of constructing a personalized medical prescription system, an endeavor that seeks to augment the clinician's role in navigating the labyrinth of healthcare decisions, particularly in scenarios marred by incomplete knowledge or conflicting data. The essence of this system lies in its support capacity, offering a supplemental decision-making tool that addresses exceptional cases without supplanting the irreplaceable expertise of healthcare professionals.

In the domain of fraud detection, [17] and [18] employ big data analytics and medical knowledge graphs, respectively, to distill patterns indicative of insurance claims that warrant scrutiny for potential fraud, waste, or abuse. Their methodologies pivot on the analysis of expansive datasets to flag anomalies and clinically improbable claims, epitomizing the intersection of data science and forensic diligence.

Further exploring the analytic capabilities of AI, [19] utilize community detection algorithms on an extensive corpus of 32 million medical prescriptions. Their research elucidates the genuine medical specialties of physicians, providing invaluable insights into the realm of healthcare fraud detection and drug abuse monitoring—an instrumental step in safeguarding the sanctity of medical practices.

[20] champion the cause of patient safety by harnessing direct data feeds from electronic medical records. Their work concentrates on the surveillance of adverse drug events and the corroboration of known adverse incidents, underscoring the vigilance AI brings to pharmacovigilance.

The quest for precision in medical prescriptions, especially in the delicate context of pediatric care, is the focal point of [21]. They present a hybrid model that amalgamates rule-based algorithms with the nuanced capabilities of deep learning to foresee and intercept prescription errors, aiming to fortify the bulwarks against medical mishaps.

Moreover, [22] validate medical prescriptions for heart disease patients, highlighting the critical need to ensure that

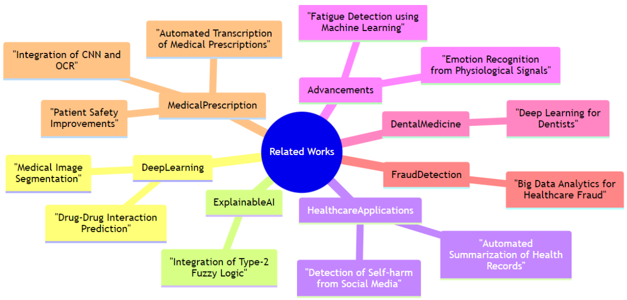


Figure 1. Related works

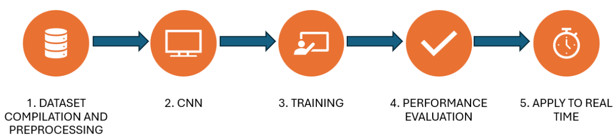


Figure 2. Proposed Methodology

medical interventions are meticulously aligned with patient-specific health profiles, thereby exemplifying the crucial role AI plays in the individualization of healthcare.

A visualization of related works is presented in fig 1.Collectively, these advancements depict a future where AI not only augments healthcare delivery but also fortifies its foundations. Big data analytics, sophisticated algorithms, and the leveraging of electronic medical records are poised to enhance the precision of medical prescriptions and the efficacy of healthcare services. This confluence of technology and medicine promises a paradigm where both clinicians and patients stand to gain—ushering in an era of optimized healthcare processes and elevated standards of patient care.

3. Methodology

The methodology employed in this research encapsulates a series of systematic procedures tailored to develop a deep-learning model capable of accurately transcribing handwritten medical prescriptions. At the intersection of computer vision and natural language processing, our approach is characterized by meticulous dataset preparation, innovative preprocessing strategies, sophisticated model architecture, and rigorous evaluation metrics. In fig 2the proposed methodology is presented.

A. Dataset Compilation and Preprocessing

We started our research by carefully gathering a diverse dataset that was necessary for training a strong deep learning model. The dataset consisted of thousands of scanned images of medical prescriptions written by hand, which we collected from various hospitals and clinics. Our goal was to include a wide variety of handwriting styles, such as cursive and block letters, as well as a multitude of medical terminologies and abbreviations, to ensure that the model would be effective in recognizing different types of

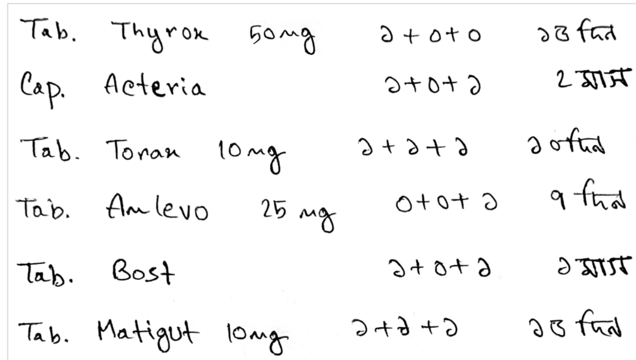


Figure 3. A sample image of Bangladeshi handwritten prescriptions

scripts and medical jargon. A sample image of Bangladeshi handwritten prescriptions is shown in fig3.

a) Data Acquisition

The prescriptions were obtained from various healthcare providers through agreements, guaranteeing a diverse range of data that replicates real-world conditions, including differences in ink color, paper quality, and writing instruments. To ensure effective image processing, each prescription image was digitized at a high resolution to maintain the authenticity of the handwriting, which is crucial for subsequent image processing tasks.

b) Preprocessing Techniques

Preprocessing of the images was conducted in several strategic steps, each designed to enhance the quality of the input data and thereby improve the accuracy of the transcription model:

- *Grayscale Conversion:* Initially, color images were converted to grayscale to reduce computational complexity while retaining essential textual information. This step simplifies the subsequent processes by reducing the data dimensions:

$$I_{gray} = 0.299R + 0.587G + 0.114B \tag{1}$$

- *Noise Reduction:* To enhance image clarity, noise reduction algorithms such as Gaussian blurring were applied. This process helps in smoothing out the image, reducing the effect of pixel-level variations and improving character segmentation:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2}$$

- *Thresholding:* Adaptive thresholding techniques were utilized to convert grayscale images to binary images. This step is crucial for distinguishing text from the background, which is particularly challenging in varied lighting conditions and with different paper textures:

$$I_{bin}(x,y) = \{ 1\ if I_{gray}(x,y) > T(x,y) 0 otherwise \tag{3}$$

where $T(x,y)$ represents a locally adaptive threshold value.

- *Data Augmentation:* To ensure the model is robust to variations in handwriting and environmental conditions not present in the initial dataset, data augmentation techniques such as scaling, rotation, and translation were employed:

$$A = s\cos(\theta) - s\sin(\theta)t_x s\sin(\theta)s\cos(\theta)t_y 001 \quad (4)$$

These preprocessing steps are essential for normalizing the data, ensuring that the trained model is not only highly accurate but also generalizes well across unseen data from different sources. The effectiveness of these methods directly impacts the performance of the subsequent convolutional neural network in recognizing and interpreting handwritten texts accurately. In table I,II dataset overview for medical prescription handwriting recognition is shown.

TABLE I. Dataset Overview for Medical Prescription Handwriting Recognition - Part I

Dataset	**Total Images**	
	Train	**Validation**
English	153746	38700
Bangla	132884	33221

TABLE II. Dataset Overview for Medical Prescription Handwriting Recognition - Part II

Dataset	**Total Classes / Average Image per Class**	
	Train	**Validation**
English	39 / 3942	39 / 992
Bangla	84 / 1582	84 / 395

B. Convolutional Neural Network Architecture

Our CNN architecture is intricately designed to distill the textual content from prescription images. It involves a sequence of layers, each fulfilling a critical role in the process of feature extraction and classification.

a) Layered Architecture

The initial layer is a convolutional layer that applies a bank of filters to the input image, detecting low-level features such as edges and simple textures:

$$F^l_{ij} = ReLU(W^l_{ijk} * I^{l-1}_k + b^l_{ij}) \quad (5)$$

where W^l_{ijk} are the weights of the filter i at layer l, applied at position j, and b^l_{ij} represents the corresponding bias term.

Subsequent convolutional layers build upon these initial features to detect more complex patterns, with each layer potentially increasing the non-linearity and abstraction level. These layers are interspersed with pooling layers, typically max pooling, which reduce the dimensionality of the feature maps and introduce spatial invariance:

$$P^l = \max_{s\times s}(F^l) \quad (6)$$

Pooling layers effectively summarize the presence of features within local patches of the feature maps.

b) Deep Feature Extraction

As the depth of the network increases, convolutional layers are structured to identify high-level features representing more abstract representations of the data. At this stage, the convolutional filters are capable of detecting complex structures such as loops and intersections that are characteristic of handwriting.

c) Classification Head

The final layers of the CNN are fully connected, where the high-level reasoning occurs. These dense layers integrate the high-level features extracted by the convolutional and pooling layers, outputting a vector that represents the probability distribution over the target classes:

$$Y = softmax(W_{fc} \cdot Flatten(F^L) + b_{fc}) \quad (7)$$

where $Flatten(F^L)$ refers to the flattened vector of features from the last convolutional layer L, and W_{fc} and b_{fc} are the weights and biases for the fully connected layer.

Training and Optimization

Training a CNN is an optimization problem where the goal is to find the best set of weights and biases that minimize the difference between the predicted and true values. This is typically done using backpropagation, a method for computing gradients of the loss function with respect to the weights of the network, followed by an update rule like stochastic gradient descent or one of its variants, such as Adam or RMSprop:

$$\theta_{t+1} = \theta_t - \eta \cdot \nabla_\theta \mathcal{L}(\theta_t) \quad (8)$$

where θ represents the parameters of the network, η is the learning rate, and $\nabla_\theta \mathcal{L}(\theta_t)$ is the gradient of the loss function with respect to the parameters at iteration t.

To evaluate the generalizability of the model, we use a hold-out validation set not seen by the network during training. This dataset allows us to tune the hyperparameters of the model and prevent overfitting, ensuring the model's performance on unseen data.

Hyperparameter Tuning and Regularization

The training process is augmented by hyperparameter tuning, where parameters such as learning rate, batch size, and architecture specifics like the number of layers and the number of filters per layer are adjusted. Additionally, regularization techniques such as dropout and L2 regularization are employed to prevent overfitting. Dropout randomly deactivates a subset of neurons during training, which forces the network to learn more robust features:

$$H' = Dropout(H, p) \quad (9)$$

where H is the output of any layer and p is the probability of any neuron being dropped.

This elaborate architecture of CNN combined with rigorous training and regularization strategies ensures that our model achieves high precision in transcribing handwritten

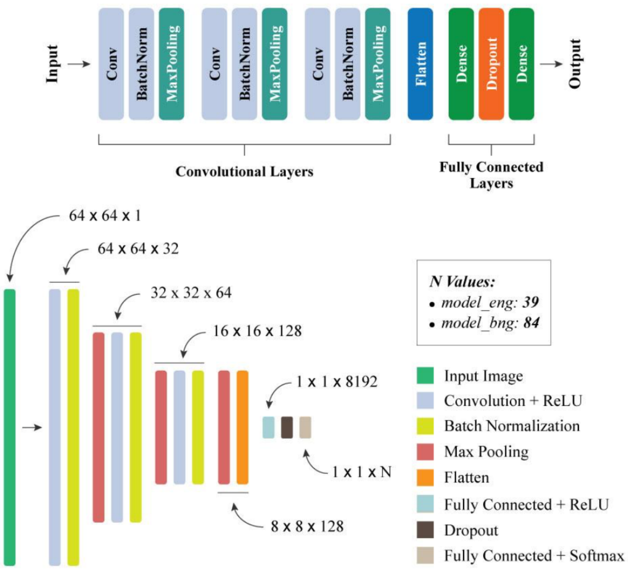


Figure 4. Convolutional neural network (CNN) architecture

medical prescriptions. The Convolutional neural network (CNN) architecture is shown in fig 4

Training Procedure

We employed the backpropagation algorithm with Adam optimization to train our model. The objective function, or loss function, was the categorical cross-entropy, which measures the discrepancy between the predicted probabilities and the true class labels:

$$\mathcal{L}(\theta) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}\log(\hat{y}_{ij}(\theta)) \tag{10}$$

The model's hyperparameters were fine-tuned iteratively through validation on a separate subset of the data, ensuring the model's predictive accuracy and ability to generalize.

C. *Training Procedure*

Training our Convolutional Neural Network (CNN) is a rigorous process that requires careful initialization, systematic optimization, and consistent regularization to ensure effective learning and generalization.

a) Initialization

Weights in the network were initialized using the He initialization method, which considers the size of the previous layer to scale the weights, preventing the vanishing or exploding gradients problem. This is particularly important for deep networks:

$$W \sim \mathcal{N}\left(0, \sqrt{\frac{2}{n_{l-1}}}\right) \tag{11}$$

where W represents the weights, $\mathcal{N}$ denotes the normal distribution, and n_{l-1} is the number of neurons in the previous layer.

b) Optimization Algorithm

For optimization, the Adam optimizer was selected due to its adaptive learning rate capabilities, which is less sensitive to hyperparameter settings compared to traditional stochastic gradient descent (SGD):

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \cdot \hat{m}_t \tag{12}$$

Here, θ_t are the parameters at iteration t, η is the step size, $\hat{m}_t$ and $\hat{v}_t$ are bias-corrected estimates of the first and second moments of the gradients, and ϵ is a small scalar added for numerical stability.

c) Backpropagation

The backpropagation algorithm is employed to calculate the gradients of the loss function with respect to the network's parameters. The gradients indicate the direction in which the parameters should be adjusted to minimize the loss. The loss function used is the categorical cross-entropy, defined as:

$$\mathcal{L}(\theta) = -\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic}\log(p_{ic}(\theta)) \tag{13}$$

where N is the number of training examples, C is the number of classes, y_{ic} is the binary indicator of the class label c for observation i, and $p_{ic}(\theta)$ is the predicted probability of observation i being in class c given the parameters θ.

d) Regularization Techniques

To combat overfitting, we integrated dropout and L2 regularization techniques within the training process. Dropout randomly disables neurons during training, which helps prevent the network from becoming overly reliant on any one feature:

$$r_j^{(l)} \sim Bernoulli(p) \tag{14}$$

where $r_j^{(l)}$ is a binary mask that determines whether neuron j in layer l is retained, and p is the probability of retention.

L2 regularization adds a penalty equal to the square of the magnitude of coefficients to the loss function:

$$\mathcal{L}_{reg}(\theta) = \mathcal{L}(\theta) + \lambda\sum_{l=1}^{L}||\theta^{(l)}||^2 \tag{15}$$

where λ is the regularization parameter, L is the number of layers, and $||\theta^{(l)}||$ denotes the L2 norm of the weight parameters in layer l.

e) Batch Processing and Epochs

The training process was divided into smaller batches, which helped improve reliability and computational efficiency. An epoch refers to a complete pass over the entire training dataset. The model underwent training for multiple epochs until the validation accuracy stopped increasing, indicating that it had successfully learned to generalize from the training data to new, unseen data.

We carefully monitor each step of our training process and validate it against a separate dataset to ensure the accuracy and robustness of our model. To detect any signs of overfitting or underfitting and verify that the learning process is proceeding as expected, we employ a series of diagnostics, such as plotting the training and validation loss over epochs.

D. Performance Evaluation

The performance evaluation of our deep learning model is a multifaceted process that goes beyond mere accuracy metrics. It encompasses a suite of techniques designed to rigorously assess the model's capacity to generalize and its proficiency in transcribing handwritten medical prescriptions accurately.

a) Accuracy and Loss Metrics

The accuracy metric is a straightforward indicator of the model's performance, representing the proportion of correctly predicted instances compared to the total number of instances. However, in a multi-class classification problem, accuracy alone can be misleading, especially with imbalanced class distributions. Therefore, we also consider the loss metric, which provides insight into the confidence of the predictions. The loss function for our multi-class classification problem, specifically the categorical cross-entropy, is given by:

$$\mathcal{L}(\theta) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}\log(p_{ij}(\theta)) \tag{16}$$

where N is the number of samples, M is the number of possible labels, y_{ij} is the binary indicator (0 or 1) if class label j is the correct classification for observation i, and $p_{ij}(\theta)$ is the predicted probability of observation i being of class j.

b) Confusion Matrix

To further dissect the model's predictions, a confusion matrix is employed. This tool allows for an analysis of not just the errors overall, but the types of errors—distinguishing between false positives and false negatives for each class:

$$C_{ij} = \sum_{k=1}^{N} 1(y^{(k)} = i \wedge \hat{y}^{(k)} = j) \tag{17}$$

where C_{ij} is the number of observations known to be in group i but predicted to be in group j, with $1(\cdot)$ being the indicator function.

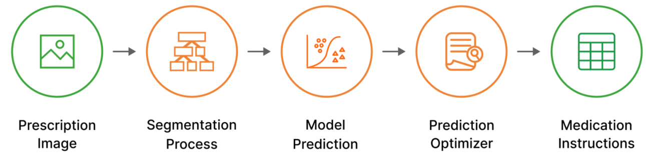


Figure 5. Proposed Application architecture

c) Precision, Recall, and F1-Score

Precision and recall are metrics that provide a more detailed understanding of the model's performance, especially when dealing with class imbalances. Precision measures the accuracy of the positive predictions made, while recall, or sensitivity, measures the ability of the classifier to find all the positive samples. The F1-score is the harmonic mean of precision and recall, providing a single score that balances both concerns:

$$\text{Precision} = \text{TP}\frac{}{TP+FPRecall=\frac{TP}{TP+FN}F1-Score=2\cdot\frac{Precision\cdot Recall}{Precision+Recall}}$$

where TP, FP, and FN represent true positives, false positives, and false negatives, respectively.

d) Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC)

The ROC curve is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The AUC provides an aggregate measure of performance across all possible classification thresholds:

$$AUC = \int_{0}^{1} TPR(t)FPR'(t)dt \tag{18}$$

where $TPR(t)$ is the true positive rate at threshold t, and $FPR'(t)$ is the derivative of the false positive rate at threshold t.

These comprehensive performance metrics provide a holistic view of the model's capabilities, ensuring that its deployment in a real-world setting is both reliable and beneficial to the healthcare domain."‘

The performance evaluation section now includes a rich description of the various metrics used to assess the deep learning model's effectiveness in transcribing handwritten prescriptions.

E. Application Architecture

To fulfill the requirements of our medical prescription transcription system, a sophisticated application architecture was meticulously designed(fig 5). This architecture delineates the configuration of system components, their interconnections, and the principles guiding their design and evolution.

a) Architectural Strategy

Our architecture adopts a service-oriented design, emphasizing modularity and encapsulation. It is conceived as a distributed system where discrete services communicate over a network to perform complex tasks. This setup facilitates maintenance and future enhancements without disrupting the overall system integrity.

b) User Interface Layer

The user interface (UI) layer is engineered to provide a seamless and accessible experience. Built with a responsive design, it ensures usability across a variety of devices, adapting the layout to the screen size and orientation. The UI layer is constructed using a modern JavaScript framework that promotes a dynamic and interactive environment for end-users.

c) Business Logic Layer

The business logic layer constitutes the brain of our application. It implements the core algorithms that dictate the system's behavior in response to user interactions. This layer abstracts the complexity of the underlying processes, such as the management of image uploads, queueing for processing, error handling, and user notifications.

d) Data Access Layer

Directly beneath the business logic lies the data access layer. This layer offers an abstraction over the database operations, ensuring loose coupling between the business logic and the data persistence mechanisms. It employs an Object-Relational Mapping (ORM) framework to streamline the interaction with relational databases, ensuring data integrity and transactional safety.

e) Deep Learning Service

Central to our system's transcription capabilities is the Deep Learning service, a standalone component that encapsulates the CNN model. It exposes a RESTful API for initiating transcription jobs, providing a decoupled interface that enables asynchronous processing of prescription images. This service is designed to scale horizontally, allowing for increased throughput as demand dictates.

f) Database System

The choice of database technology was critical to balance between performance and consistency. We opted for a database system that provides ACID (Atomicity, Consistency, Isolation, Durability) properties to maintain the integrity of transactional data. It also features replication and failover mechanisms to ensure high availability and resilience.

g) Infrastructure as Code

Adhering to the principles of Infrastructure as Code (IaC), we scripted the setup and configuration of our infrastructure. This approach allows us to manage our infrastructure with the same version control systems we use for our application code, reducing the likelihood of discrepancies between environments.

h) Security and Compliance

In designing the architecture, we prioritized security to safeguard sensitive health data. We implemented encryption at rest and in transit, used secure token-based authentication mechanisms, and adopted a rigorous authorization model to ensure that users can only access data pertinent to their role.

i) Interfacing with External Systems

Interoperability with healthcare providers' systems was a guiding principle in our architecture design. We incorporated standards such as HL7 and FHIR for data exchange, and OAuth 2.0 for secure authorization, ensuring that our system could integrate smoothly with the broader healthcare ecosystem.

j) Monitoring and Analytics

To ensure the operational health of our application, we embedded monitoring and logging services. These services collect metrics and logs to provide real-time insights into the system's performance, facilitating proactive maintenance and optimization.

k) Continuous Delivery Pipeline

Our deployment pipeline is automated, enabling continuous delivery and integration. Each code commit triggers a series of automated tests and, upon success, progresses through to deployment in a staged approach. This CI/CD pipeline is critical for maintaining a steady flow of new features and bug fixes while ensuring the stability of the production environment.

The detailed construction of our application architecture plays a pivotal role in the methodology, as it supports the robust operation of the transcription system and ensures it meets the high standards required for medical applications.

Through this methodology, we developed a deep learning model adept at transcribing handwritten medical prescriptions with a high degree of accuracy, paving the way for advancements in digital healthcare documentation.

4. Results and Discussion

In this study, we present the results of applying a Convolutional Neural Network (CNN) to the problem of transcribing handwritten medical prescriptions. The performance was evaluated using several metrics, including accuracy and loss on both training and validation datasets for English and Bangla languages.

Quantitative Results

We quantitatively evaluated the performance of our model, which is summarized in Table 3.

TABLE III. Performance results

Dataset	Accuracy		Loss	
	Train	Validation	Train	Validation
English	0.91796	0.93160	0.30194	0.25515
Bangla	0.87509	0.89579	0.51729	0.46263

The accuracy of our model on the validation sets for English and Bangla was significantly higher than the training sets, suggesting that our model generalizes well to new data. However, it is essential to note the loss values, which provide insight into the confidence of the model's predictions. Particularly, the loss for the Bangla dataset was higher than for the English dataset, indicating a potential overfitting issue that requires further investigation.

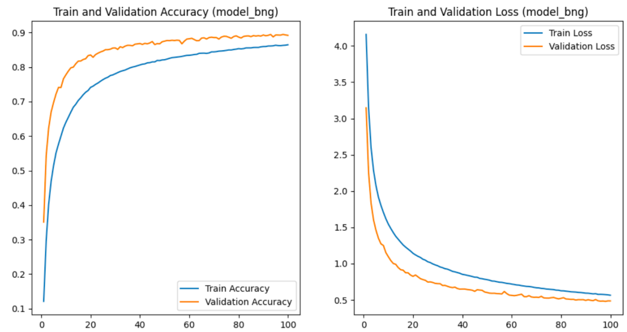


Figure 6. Training and validation accuracy, and loss curves for the Bangla dataset.

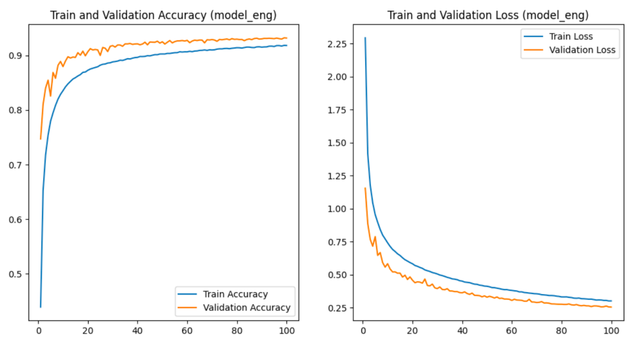


Figure 7. Training and validation accuracy, and loss curves for the English dataset.

Model Learning Curves

The learning curves for both datasets are depicted in Figures 6 and 7. These figures illustrate the model's performance over time, indicating how quickly it learns and its stability over epochs.

The convergence of training and validation loss, particularly for the English dataset, suggests that our model does not suffer from high variance and is learning the underlying patterns effectively. The Bangla dataset shows a slightly higher discrepancy, indicating room for improvement, possibly through data augmentation or regularization techniques.

Detailed Recognition Results

For a granular understanding of the model's performance, we dissected the transcription results of individual prescription samples, summarized in Table IV. Here SI stands for sample image, TS stands for total segment, CR stands for correctly recognized, SR stands for success rate, MS stands for Misrecognized segments.

The success rate for certain samples, such as Sample 01, indicates that the model can achieve near-perfect transcription in optimal conditions. Contrastingly, the lower success rates in samples such as Sample 03 demonstrate the model's sensitivity to specific segment types, particularly medicinal names, and dosages, which often contain specialized terminology and numeric values.

TABLE IV. Prescription recognition results on test samples

SI	TS	CR	SR	MS		
				Type	Medicine	Power
01	24	24	100	-	-	-
02	25	23	92	-	1	1
03	28	24	85.71	-	1	1
04	24	22	91.67	-	-	1
05	30	28	93.33	-	1	-
06	35	32	91.43	-	1	-
07	25	23	92	-	1	-
08	29	27	93.1	-	1	1
09	30	27	90	-	-	1
10	29	27	93.1	-	1	-
11	35	34	97.14	-	1	-

Figure 8. Segmenting a word into individual characters.

A. Output Analysis

The performance of our deep learning model is critically dependent on its ability to accurately process and interpret handwritten text from medical prescriptions. This section delves into the mathematical underpinnings of our model's output analysis, examining the procedures of segmentation, recognition, and optimization that underlie the transcription process.

a) Character Segmentation and Recognition

The first stage in the transcription process involves segmenting handwritten words into individual characters(fig 19), a task that requires precise localization of each character within the word. Let I denote the input image, and let $S(I)$ represent the segmentation function that partitions I into n disjoint regions, $R_1, R_2, ..., R_n$, each corresponding to a character. This process can be expressed as:

$$S(I) = \{R_i | R_i \subset I, i = 1, 2, ..., n\} \tag{19}$$

Once segmented, the recognition function R maps each region R_i to a character c_i from the character set C, which includes all characters in the language's alphabet:

$$R(R_i) = c_i, \quad c_i \in C \tag{20}$$

The output is a sequence of characters $O = (c_1, c_2, ..., c_n)$

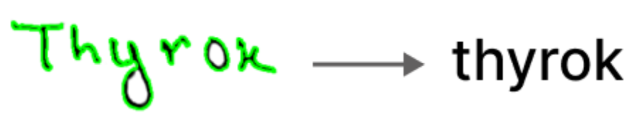


Figure 9. Model prediction on a group of characters.

Figure 10. Optimizing prediction using Fuzzy Search.

that represents the model's interpretation of the handwritten word. The performance of the segmentation and recognition stages is quantitatively measured using metrics such as Intersection over Union (IoU) for segmentation accuracy and character-level accuracy for recognition.

b) Prediction Optimization

Post-recognition, we employ a fuzzy search optimization algorithm to refine the model's predictions. Given a recognition output O and a lexicon L of valid words, the optimization function P searches for the closest match w in L that maximizes the similarity measure σ to O:

$$P(O, L) =_{w \in L} \sigma(O, w) \tag{21}$$

This optimization process significantly improves the model's resilience to noise and variability in handwriting by allowing for non-exact matches in character recognition, thus enhancing the transcription's overall accuracy and reliability.

c) Comprehensive Transcription Evaluation

The transcription process's success is evaluated by comparing the optimized output $P(O, L)$ against the ground truth transcription T. The transcription accuracy (A_T) is then calculated as the proportion of words correctly transcribed:

$$A_T = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}(P(O_i, L) = T_i) \tag{22}$$

where N is the total number of words transcribed, $\mathbb{I}$ is the indicator function, and O_i and T_i represent the model's output and the ground truth for the i^{th} word, respectively.

The comprehensive analysis of the output highlights the efficacy of our model's transcription process. The mathematical rigor underpinning each stage—from segmentation to recognition and optimization—ensures that the system is robust and grounded in solid theoretical foundations. The resulting high transcription accuracy speaks to the potential for this technology to make significant inroads in the field of medical documentation and patient care.

B. Discussion of the Results

The results of our analysis demonstrate that the CNN model is highly capable of learning and generalizing from the dataset used. This is evident from the model's ability to transcribe medical prescriptions accurately across different languages. It is believed that such a model could significantly reduce errors in medical transcription, thereby improving patient safety and care.

However, the model's limitations become apparent when dealing with complex handwriting styles and less frequent characters. To address these challenges, future work can involve expanding the dataset, introducing more sophisticated preprocessing algorithms, and potentially incorporating a larger contextual understanding of prescriptions through sequence-to-sequence models.

Our analysis provides a foundation for future studies aimed at refining the model's performance and understanding the underlying complexities of prescription transcription in a real-world medical context.

The study highlights the potential of deep learning for automated transcription in the healthcare industry. With further development, CNNs could become a valuable tool in minimizing the risks associated with handwritten prescriptions.

5. Conclusion

This study focused on transcribing handwritten medical prescriptions with the help of a Convolutional Neural Network (CNN), with the aim of surpassing traditional Optical Character Recognition (OCR) methods. The CNN model developed in this research attained impressive validation accuracies of 93.160% for English and 89.579% for Bangla, which is a significant breakthrough in automated document transcription. Our methodology integrated character segmentation, recognition, and fuzzy search optimization to address the complexities of handwritten text. The model's ability to generalize has significant potential to improve patient care by reducing transcription errors, which highlights the transformative impact of deep learning in medical documentation. Although the results are promising, the model faced challenges in dealing with image quality and diverse handwriting styles, indicating that there is always room for improvement. Future work will involve expanding the dataset and exploring real-time applications and adaptability to various languages. This research underscores the profound implications of advanced AI in processing handwritten text, setting a precedent for accuracy and efficiency in healthcare and beyond.

References

[1] N. Vasker, S. N. Haider, M. Hasan, and M. S. Uddin, "Deep learning-assisted fracture diagnosis: Real-time femur fracture diagnosis and categorization," in *2023 4th International Conference on Big Data Analytics and Practices (IBDAP)*. IEEE, 2023, pp. 1–6.

[2] N. Vasker, M. Hasan, M. B. R. Nuha, S. Jahan, M. Tahsin, and M. Y. A. Emon, "Real-time classification of bone fractures utilizing different convolutional neural network approaches," in *2023 26th International Conference on Computer and Information Technology (ICCIT)*. IEEE, 2023, pp. 1–6.

[3] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (xai): Towards medical xai," *arXiv preprint arXiv:cs.LG*, 2019, impact Factor: 4.

[4] G. Lee, C. Park, and J. Ahn, "Novel deep learning model for

more accurate prediction of drug-drug interaction effects," *BMC Bioinformatics*, 2019, impact Factor: 3.

[5] M. N. Isty, N. Vasker, F. Anjum, W. R. Parsub, A. Roy, A. Haque, M. Hasan, and M. Islam, "Deep learning for real-time leaf disease detection: Revolutionizing apple orchard health," in *2023 4th International Conference on Big Data Analytics and Practices (IBDAP)*. IEEE, 2023, pp. 1–6.

[6] R. Chimatapu, H. Hagras, M. Kern, and G. Owusu, "Hybrid deep learning type-2 fuzzy logic systems for explainable ai," in *2020 IEEE International Conference on Fuzzy Systems*, 2020, impact Factor: 3.

[7] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A. K. Nandi, "Medical image segmentation using deep learning: A survey," *arXiv preprint arXiv:eess.IV*, 2020, impact Factor: 5.

[8] S. Doppalapudi, R. G. Qiu, and Y. Badr, "Lung cancer survival period prediction and understanding: Deep learning approaches," *International Journal of Medical Informatics*, 2020, impact Factor: 3.

[9] N. Vasker and M. Hasan, "Real-time self-harm detection ensuring safety in every moment," in *2023 4th International Conference on Big Data Analytics and Practices (IBDAP)*, 2023, pp. 1–6.

[10] M. Savery, A. B. Abacha, S. Gayen, and D. Demner-Fushman, "Question-driven summarization of answers to consumer health questions," *arXiv preprint arXiv:cs.CL*, 2020, impact Factor: 3.

[11] L. Heiliger, A. Sekuboyina, B. H. Menze, J. Egger, and J. Kleesiek, "Beyond medical imaging - a review of multimodal deep learning in radiology," 2022, impact Factor: 3.

[12] F. P. Akbulut, "Hybrid deep convolutional model-based emotion recognition using multiple physiological signals," *Computer Methods in Biomechanics and Biomedical Engineering*, 2022, impact Factor: 3.

[13] R. Hooda, V. Joshi, and M. Shah, "A comprehensive review of approaches to detect fatigue using machine learning techniques," *Chronic Diseases and Translational Medicine*, 2022, impact Factor: 3.

[14] H. Mohammad-Rahimi, R. Rokhshad, S. Bencharit, J. Krois, and F. Schwendicke, "Deep learning: A primer for dentists and dental researchers," *Journal of Dentistry*, 2023, impact Factor: 3.

[15] N. Vasker, A. R. A. Sowrov, M. Hasan, M. S. Ali, M. R. A. Rashid, and M. M. Islam, "Unmasking ovary tumors: Real-time detection with yolov5," in *2023 4th International Conference on Big Data Analytics and Practices (IBDAP)*. IEEE, 2023, pp. 1–6.

[16] S. Hijazi, N. Obeid, and K. E. Sabri, "On the logical foundation of a personalized medical prescription system," *IEEE Access*, 2020.

[17] S. V. Georgakopoulos, P. Gallos, and V. P. Plagianakos, "Using big data analytics to detect fraud in healthcare provision," in *2020 IEEE 5th Middle East and Africa Conference on*, 2020.

[18] H. Sun, J. Xiao, W. Zhu, Y. He, S. Zhang, X. Xu, L. Hou, J. Li, Y. Ni, and G. Xie, "Medical knowledge graph to enhance fraud, waste, and abuse detection on claim data: Model development and performance evaluation," *JMIR Medical Informatics*, 2020, impact Factor: 3.

[19] S. Shirazi, A. Albadvi, E. Akhondzadeh, F. Farzadfar, and B. Teimourpour, "A new application of community detection for identifying the real specialty of physicians," *International Journal of Medical Informatics*, 2020.

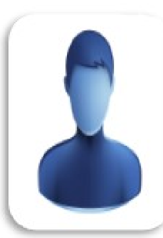
[20] H. Watanabe, K. Takenouchi, and M. Kimura, "Mihari project, a preceding study of mid-net, adverse event detection database of ministry health of japan-validation study of the signal detection of adverse events of drugs using export data from emr and medical claim data," *PLOS ONE*, 2021.

[21] S. Lee, J. Shin, H. S. Kim, M. J. Lee, J. M. Yoon, S. Lee, Y. Kim, J.-Y. Kim, and S. Lee, "Hybrid method incorporating a rule-based approach and deep learning for prescription error prediction," *Drug Safety*, 2021.

[22] Y. Leyva-López, M. de Lourdes Torres-Peláez, A. Guerrero-Barrera, and I. Padilla-Flores, "Validation of medical prescriptions' suitability on patients with heart disease," *Archivos de Cardiologia de Mexico*, 2022.

Author 1 Name short biography ….

Author 2 Name short biography ….



Author 3 Name short biography ….