IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 1, January 2025

DOI: 10.17148/IARJSET.2025.12107

Enhancing Patient Safety: A Hybrid CNN- BiLSTM Approach for Analysis of Doctor's Handwritten Prescriptions

Swarnalata Bollavarapu[1], Nilesh S. Bhelkar[2], Kshitiz Sharma[3], Neel Chopra[3], Yash Korla[3]

Assistant Professor, Computer Engineering Department, Mukesh Patel School of Technology Management and Engineering, NMIMS University, Mumbai, INDIA[1]

Assistant Professor, Artificial Intelligence and Data Science Department,

Rajiv Gandhi Institute of Technology, Mumbai, INDIA[2]

Student, Computer Engineering Department, Mukesh Patel School of Technology Management and Engineering, NMIMS University, Mumbai, INDIA[3]

Abstract: Accurate interpretation of doctor's handwritten prescription is a critical task in healthcare for the patient safety and minimization of medication errors. Moreover, illegible handwriting poses a great risk, which leads to misinterpretation resulting in adverse health outcomes in many cases. Recent breakthroughs in AI, more so the deep learning segment, introduce robust methods to automate handwritten text interpretations, hence providing a way to streamline processes and improve the accuracy in the healthcare setup. In previous studies, models like CNN, RNN were explored, but the standalone CNN and RNN were giving suboptimal responses. In this study, a Hybrid CNN-BiLSTM model is proposed to predict and analyze words in handwritten prescriptions. This research utilizes CNN for capturing the spatial features and BiLSTM networks for analyzing sequential dependencies, which makes this approach suitable for complex handwriting pattern recognition in medical documents. To evaluate the performance of the proposed model, its performance was compared with Google's Vision API. It is a machine learning-powered service for image content analysis. The results are indicative of the great potential that lies in the application of the CNN-BiLSTM architecture for advancements in automated prescription analysis to improve patient safety and operational efficiency within healthcare settings.

Keywords: Convolutional Neural Network (CNN), Bidirectional Long-Short Term Memory (Bi-LSTM), Deep Learning, Machine Learning

I. INTRODUCTION

In the recent past, Artificial Intelligence (AI) has led to rapid advancements in the healthcare sector with the aim of improving patient safety. One such area is of Doctor's handwritten prescription analysis. Illegible writing in prescriptions leads in wrong interpretation of medicine or sometimes dosage of medicines, thereby impacting the patient health. Recognition of medical text from images automatically and accurately is a crucial factor that contributes towards betterment in health efficiency as well as in patient care.

Based on the literature review carried out, it is observed that techniques such as Optical Character Recognition (OCR), ML and Natural Language Processing (NLP) techniques have been employed. Previous work has emphasized the importance of accurate prescription analysis to prevent errors and the need for enhancing the work by automating the medication identification, dosage calculations and drug interaction checks.

This study investigates the recognition of medicine names from Prescription images by incorporating a Convolutional Neural Network with a Bidirectional Long Short-Term Memory architecture. CNNs are good at handling spatial features in images, while Bi-LSTMs have great performance in modelling sequences; hence, for complicated text recognition in medical images, this combination is ideal.

It evaluates the performance of this hybrid architecture by comparing its performance with Google's Vision API. It is one of the most widely used cloud-based services used for image analysis. The mainstay of text extraction from pictures in Google Vision API is a set of deep learning models that have been trained on a large dataset of images that have set this

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⩱ Peer-reviewed / Refereed journal ⩱ Vol. 12, Issue 1, January 2025

DOI: 10.17148/IARJSET.2025.12107

tool as the standard in the area of OCR. The following performance metrics such as accuracy, precision, recall is used in the study. For better user experience, a graphical user interface (GUI) is also built.

The contribution in this paper lays the foundation for future advances in medical image analysis by exploring trade-offs involving bespoke models and established solutions like the Google Vision API. It will also serve as an exceedingly coherent path toward fine-tuning for improvements on image pre-processing, character recognition, and scalability of the system for better applicability in real-world healthcare AI-driven solutions.

The rest of the paper is organized as follows: Section II describes the research and development relevant to this work. Section III describes the proposed approach, its theoretical framework, and expected benefits. Section IV deals with Implementation and Results, where the details of practical implementation and assessment of the results are presented. Finally, Section V presents the summary of the paper and suggest directions of further research.

II. LITERATURE REVIEW

Concomitantly, the rate and complexity increase the potential adverse health outcome of a mistake. The review [1] has discussed various AI methodologies applied to selected tasks, including NLP, Machine Learning, and Deep Learning-complete medication identification, dosage calculation, and checks against drug interaction. It further looks at the current applications of AI, presents some case studies on improved accuracy and efficiency of prescriptions, and acknowledges challenges in relation to data quality and ethical considerations.

The survey classifies the different HTR systems according to techniques, datasets, and levels of recognition, analyzing also the accuracy of the latter. It observes that these recurrent layers perform much better than the MDLSTM models and underlines the efficiency of data augmentation on systems trained with the RIMES database. Finally, it gives a comparison of commercial HTR systems such as Ocelus, Transkribus, and DOCSUMO; their English datasets are more accurate [2].

Medical data mining is becoming more and more important with the rapid increase in electronic publications, which are difficult for traditional text-mining tools to process. That has resulted in the application of AI-based tools that perform well when it comes to analysing big datasets-reaching out to their hidden patterns and improving accuracy. The review provides an overview of the CRISP-DM process, challenges such as heterogeneity and quality in data mining of biomedical data, along with ethical or legal concerns. Also, it describes the data extraction carried out by web scraping tools and the involvement of information management systems within health care. Text mining techniques will be essential to gain insights; however, having proper data preparation and modeling would help visualize and effectively communicate the findings [3].

The paper reviews the revolutionary impact brought by deep learning in natural language processing, underlining major advances that have so far been achieved and challenges yet to be overcome. Representation learning with transformers, transfer learning with models such as BERT and GPT, sequence-to-sequence models, attention mechanisms, and multimodal NLP are considered the main developments. Nevertheless, challenges remain with large annotated datasets, a lack of interpretability, domain adaptation, ethical concerns regarding bias, and verification of robustness in diverse settings. The paper thus discusses ways in which these challenges need to be overcome for the responsible deployment and further development of deep learning in NLP and explores future research directions in making models more interpretable, addressing bias, and developing more efficient models for low-resource settings [4].

In this article, applications of AI have been discussed in clinical decision-making, image analysis, and target discovery. It covers some of the critical applications of NLP in biomedicine: from protein-protein interaction prediction to disease diagnosis; a number of techniques for natural language processing, such as word embeddings and neural networks; biomedical summarization challenges; and the need for XAI to help AI be interpretable and trustworthy by focusing on model decisions understanding health care [5].

The methodologies, challenges, and advancements of the HWR are reviewed in this article, focusing on a comprehensive survey by machine learning algorithms. Basically, HWR can be understood whereby the computer is able to detect and recognize human handwriting input from different sources such as paper documents and digital interfaces. This study further distinguishes between online and offline handwriting recognition and gives weight to the aspect of digit identification in banking and document processing. Most of the machine learning techniques, including CNNs explored

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⵣ Peer-reviewed / Refereed journal ⵣ Vol. 12, Issue 1, January 2025

DOI: 10.17148/IARJSET.2025.12107

in this review, are effective in improving recognition accuracy. It also presents the future course of research and applications toward the improvement of accuracy and the scope for expanded recognition systems [6].

The present paper, proposes the enhancement of handwriting recognition systems with Temporal Dropout-TD, a technique designed for Recurrent Neural Networks. The proposition hence shall be made to try and overcome the well-known problem of overfitting in long short-term memory networks by randomly dropping information at various places within input sequences during training. Here, the authors apply TD both at the level of image data and the internal network representations. Their experiments show that TD significantly enhances performance on handwriting recognition and improves over previous state-of-the-art methods [7].

The authors have compared the performance of the CNN with an SNN and show that the CNN significantly outperforms the SNN in both accuracy and speed. The presented state-of-the-art discusses increasing the size of a dataset, optimal epoch choices when training, and hardware enhancement for improved recognition. The emphasis will be more on how this technology may be used in different sectors while at the same time pointing out its limitations with regard to overcoming the various hurdles for commercial application [8].

This paper's proposed system enhances previously proposed models through the incorporation of real-life prescription data and tuning of hyperparameters for improved accuracy. It proposes a solution with a reduced character error rate and word error rate compared to the conventional methods. The Lexicon Search algorithm greatly improves the word accuracy by matching predictions against a deep database of drug names [9].

This article develops a web application, "Mediscan", to interpret the handwriting of doctors in prescriptions and offer the best price of medicines across various e-commerce platforms. Mediscan incorporates OCR and NLP techniques for correctly extracting the names of medicines from scanned prescriptions. Then, the application compares the price from sites like 1Mg, Truemeds, and Amazon [10].

The study "Doctor's Cursive Handwriting Recognition System Using Deep Learning" had proposed a system to recognize and translate the often-illegible cursive handwriting of doctors into readable text. A Handwriting Recognition System using Deep Convolutional Recurrent Neural Network has been developed for translating handwritten medical prescriptions to machine-readable text [11].

This research paper, "Doctor's Handwritten Prescription Recognition System in Multi-Language Using Deep Learning," puts forth a solution to interpret the mostly illegible handwritten prescriptions of doctors that may lead to medication errors later on. The system applies deep learning techniques such as CNN, RNN, and LSTM in recognizing and thereby converting handwritten prescriptions into readable digital text. The use of OCR, fuzzy search, and market basket analysis improves the accuracy and usability of prescription recognition in numerous languages [12].
This paper provides a wide-ranging review of deep learning model developments used in Natural Language Processing. It traces the transition from early statistical models such as HMMs and CRFs to neural networks such as RNNs, LSTMs, and finally transformer models like BERT and GPT [13].
Table 1 presents a summary of notable work in terms of the datasets employed and the obtained accuracy results.

Table 1: Summary of datasets used

Paper	Dataset Used	Accuracy (%)
Zia, A., et al. (2022)	MIMIC-III and other open medical datasets	Varies widely, but can reach up to 90% for specific tasks
Chammas, E., & Mokbel, C. (2021)	IAM Handwriting Database	Approximately 85-90%
Razdan, A., et al. (2023)	Self-collected handwritten prescription dataset	Up to 92%
Joshi, T., et al. (2023)	Proprietary dataset of prescription images	Approximately 88%
Fajardo, L.J., et al. (2019)	Self-collected handwritten prescription dataset	Around 87%
Pavithiran, G., et al. (2022)	dataset of multi-language prescriptions	About 89%

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 1, January 2025

DOI: 10.17148/IARJSET.2025.12107

III. PROPOSED WORK

The following sections presents details of the proposed methodology and figure 1 shows the flow of the project.

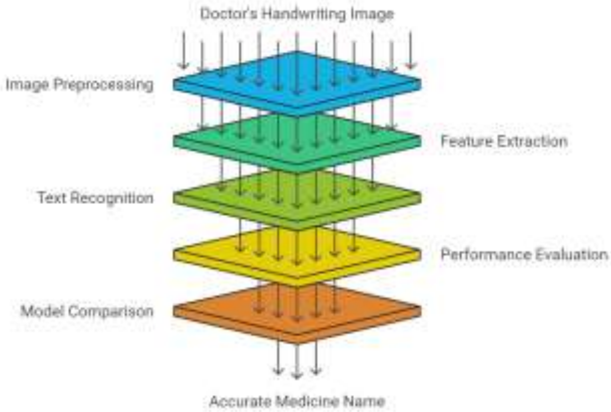


Figure 1. Flow of the project

Step 1-Image Pre-processing:

i. Image Segmentation: The image is segmented based on the “spacing” between the words.

ii. Resizing the Image: The image is resized to a standardized dimension (32x128) to ensure uniformity and make it suitable for input into the model.

iii. Casting Image to tf.float32: The resized image is then converted to a tf.float32 data type to ensure compatibility with TensorFlow operations during model training and inference.

iv. The output of this step is a processed image, ready for use in the subsequent recognition model.

Figure 2 shows the visual representation of pre-processing steps

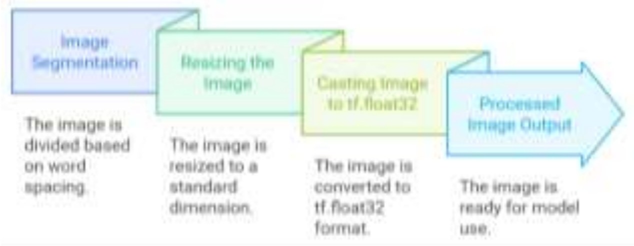


Figure 2 Image Pre-Processing steps

The experimented work consisted of a Bidirectional Long Short-Term Memory (Bi-LSTM) network, structured with specific layers designed to optimize performance.

Step 2- Medicine Name Recognition:

i. CNN: Extracts spatial features from the image, capturing patterns such as edges and shapes, which are crucial for recognizing characters and text structure.

ii. Bi-LSTM: Processes the spatial features in both forward and backward directions, allowing the model to understand dependencies and improve text recognition accuracy.

iii. The output of this stage is the recognized medicine name, predicted by the CNN + Bi-LSTM model.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 1, January 2025

DOI: 10.17148/IARJSET.2025.12107

Figure 3 shows the model architecture.

```
Layer (type)                Output Shape         Param #    Connected to
==========================================================================
image (InputLayer)          [(None, 128, 32, 1)  0          []
                            ]

Conv1 (Conv2D)              (None, 128, 32, 32)  320        ['image[0][0]']

Conv4 (Conv2D)              (None, 128, 32, 128  36992      ['Conv1[0][0]']
                            )

pool1 (MaxPooling2D)        (None, 64, 16, 128)  0          ['Conv4[0][0]']

Conv6 (Conv2D)              (None, 64, 16, 256)  295168     ['pool1[0][0]']

dropout (Dropout)           (None, 64, 16, 256)  0          ['Conv6[0][0]']

Conv8 (Conv2D)              (None, 64, 16, 1024  2360320    ['dropout[0][0]']
                            )

pool2 (MaxPooling2D)        (None, 32, 8, 1024)  0          ['Conv8[0][0]']

Conv7 (Conv2D)              (None, 32, 8, 64)    589888     ['pool2[0][0]']

...
Total params: 23,294,481
Trainable params: 23,294,481
Non-trainable params: 0
```

Figure3 Model Architecture

Step 3- Validation:

i. Accuracy: The proportion of correctly identified medicine names, indicating overall model performance. The accuracy was calculated using Mean Edit Distance (MED) Metric.

ii. Loss: The error calculated during model training using a specific loss function (such as Connectionist Temporal Classification loss), which guides model optimization.

Step 4- Comparison with Google Vision API:

i. To compare the model's performance with benchmarked Google's Vision API was utilized.

ii. The outputs from the proposed work were compared against those from the Vision API to assess the relative strengths and weaknesses of this approach in terms of accuracy and recognition capabilities.

IV. IMPLEMENTATION AND RESULTS

The system is developed using Python, TensorFlow, Keras, Streamlit and Google's Vision API used for comparison. The dataset used in this work consisted of around 4,800 images. The dataset was a combination of IAM handwriting database [14] from Kaggle and handwritten prescriptions collected from a doctor. This model achieved a maximum accuracy of 79.5%

To assess the model's performance, the edit distance was implemented—a measure that calculates a minimum number of operations capable of transforming one string into another, the formula of which has been depicted in Equation 1 and Equation 2. During the training, CTC loss was utilized, which is fairly effective in pattern recognition problems with sequences of variable lengths, like speech or handwriting.

$$D(i,j) = \begin{cases} 0 & \text{if } i = 0 \text{ and } j = 0, \\ j & \text{if } i = 0 \text{ (all insertions)}, \\ i & \text{if } j = 0 \text{ (all deletions)}, \\ \min \begin{cases} D(i-1,j)+1 & \text{(deletion)}, \\ D(i,j-1)+1 & \text{(insertion)}, \\ D(i-1,j-1)+c(a_i,b_j) & \text{(substitution)} \end{cases} & \text{if } i,j > 0, \end{cases}$$

where $c(a_i, b_j) = 0$ if $a_i = b_j$ (no cost for matching characters) and $c(a_i, b_j) = 1$ if $a_i \neq b_j$ (substitution cost).

Eq.1 Mathematical Representation of Edit Distance

$$MED = \frac{1}{N} \sum_{k=1}^{N} D\ (label_k, prediction_k), \ldots\ldots.. \text{ Eq.2}$$

Where N is the number of label-prediction pairs in the batch

Figure 4 shows the output of the image pre-processing steps on a sample image having a text "And Philip said".

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ☵ Peer-reviewed / Refereed journal ☵ Vol. 12, Issue 1, January 2025

DOI: 10.17148/IARJSET.2025.12107

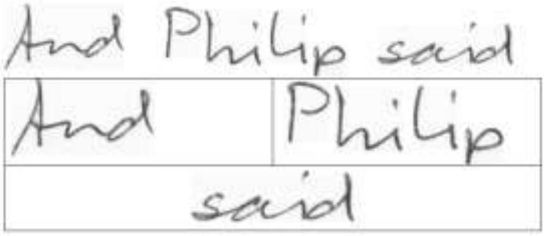


Fig 4 Output of Pre-processing step

Below figure 5, 6 and 7 shows results obtained of proposed work along with google vision API results. The results show the correctly recognised words from the three sample image inputs.

.Input:

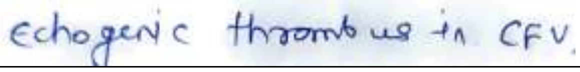
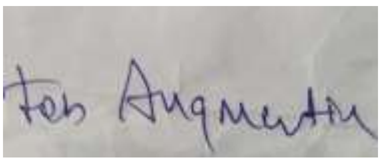


Figure.5 Sample Input 1 and its Output Screenshot

The above output shows that the proposed model is able to detect 70-80% of the letters correctly, while maintaining the relative order of the words; whereas; Google's model falls short to maintain the order.

.Input: 

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 1, January 2025
DOI: 10.17148/IARJSET.2025.12107

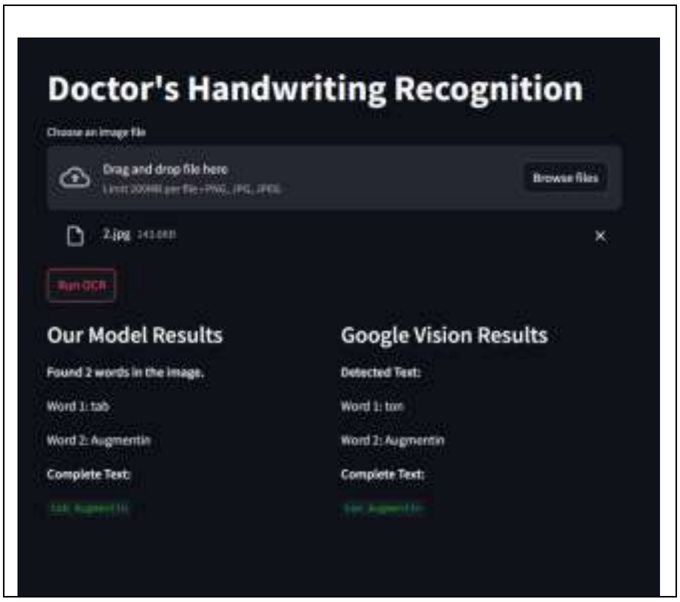


Figure 6 Sample Input 2 and its Output Screenshot

The above output shows that the proposed model was correctly able to predict “tab”, whereas Google’s model predicted it as “ton”.

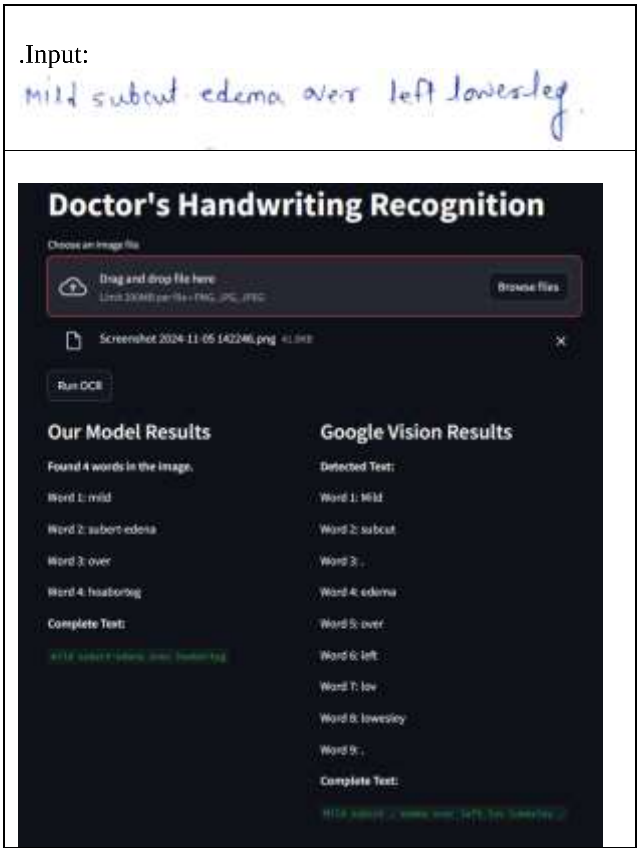


Figure 7 Sample Input 3 and its Output Screenshot

The above output shows that both models are unable to predict the words correctly, and the proposed model lacks the knowledge of special characters, and ignores them.

DOI: 10.17148/IARJSET.2025.12107

V. CONCLUSION AND FUTURE WORK

A. CONCLUSION

This project contributes a robust system for recognizing medicine names from images using a combination of CNN and Bi-LSTM architectures. Comparing the model's outputs with those of Google Vision API further validates the effectiveness of the approach. Anticipated future improvements, such as enhancing special character recognition, refining preprocessing for blurry images, and extending model training, will further boost the system's accuracy and applicability. This research lays a solid foundation for developing intelligent systems capable of reliable text recognition in medical and pharmaceutical contexts.

B. FUTURE SCOPE

Enhancement of Special Character Recognition – The model will be refined to improve its ability to recognize special characters, which is crucial when medicine names contain symbols or punctuation marks commonly used in pharmaceutical labelling.

Optimizing Image Preprocessing – Current preprocessing techniques will be improved to handle blurry or distorted images more effectively. Advanced image processing methods, such as deblurring algorithms and adaptive contrast adjustments, will be explored to ensure clearer and more consistent input, thus improving recognition accuracy.

Extended Model Training – The model will undergo additional training with an expanded and more diverse dataset. This will help the model generalize better and improve its performance across different image conditions and text variations.

REFERENCES

[1] Agrawal, M., Mishra, V., Jogani, H.S., Kashyap, P. and Mishra, R.G., 2023, June. A Review of Artificial Intelligence in Medical Prescription Analysis. In 2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN) (pp. 35-43). IEEE.
[2] AlKendi, W., Gechter, F., Heyberger, L. and Guyeux, C., 2024. Advancements and challenges in handwritten text recognition: A comprehensive survey. Journal of Imaging, 10(1), p.18.
[3] Zia, A., Aziz, M., Popa, I., Khan, S.A., Hamedani, A.F. and Asif, A.R., 2022. Artificial intelligence-based medical data mining. Journal of Personalized Medicine, 12(9), p.1359.
[4] Deep Learning Approaches for Natural Language Processing Advancements and Challenges
[5] Moradi, M. and Samwald, M., 2022. Deep learning, natural language processing, and explainable artificial intelligence in the biomedical domain. arXiv preprint arXiv:2202.12678.
[6] Prasantha, H.S., Detailed Survey Of Handwriting Recognition Using Machine Learning Algorithms.
[7] Chammas, E. and Mokbel, C., 2021. Fine-tuning Handwriting Recognition systems with Temporal Dropout. arXiv preprint arXiv:2102.00511.
[8] Handwritten Character Recognition Using Deep Learning (Convolutional Neural Network).
[9] Razdan, A., Raghavan, A., Panandikar, M., Pol, A., Hedaoo, K. and Patil, R., 2023, July. Recognition of Handwritten Medical Prescription using CNN Bi-LSTM With Lexicon Search. In 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-6). IEEE.
[10] Joshi, T., Das, S., Ghosh, S., Jha, C. and Shinde, V.D., 2023. Web Application for Interpretation of Doctor's Handwritten Prescription and Suggesting the best Price Offer over Various e-Commerce Websites using AI.
[11] Fajardo, L.J., Sorillo, N.J., Garlit, J., Tomines, C.D., Abisado, M.B., Imperial, J.M.R., Rodriguez, R.L. and Fabito, B.S., 2019, November. 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) (pp. 1-6). IEEE.
[12] Pavithiran, G., Padmanabhan, S., Divya, N., Aswathy, V. and Chandar, B., 2022, December. Doctor's Handwritten Prescription Recognition System In Multi-Language Using Deep Learning. In 2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS) (pp. 1-5). IEEE.
[13] Chai, Y., Jin, L., Feng, S. and Xin, Z., 2024. Evolution and advancements in deep learning models for Natural Language Processing. Applied and Computational Engineering, 77, pp.144-149.
[14] IAM Handwriting database, https://www.kaggle.com/datasets/nibinv23/iam-handwriting-word-database

