International Journal of Computer Techniques

RESEARCH ARTICLE OPEN ACCESS

Transforming Handwriting Recognition: Comparative Analysis of CNN-BiLSTM and ViT-Transformer Encoder Architectures

Manmohan Jatav, M. Tech Research Scholar*, Shivank Soni, Assistant Professor**
*(Department Of Computer Science & Engineering, Oriental Institute of Science & Technology, Bhopal, MP, India
Email: manmohanj8268@gmail.com)
** (Department Of Computer Science & Engineering, Oriental Institute of Science & Technology, Bhopal, MP, India
Email : shivanksoni@gmail.com)

--************************----------------------------------

Abstract:

This study undertakes a somewhat comparative analysis of the transformation of the offline handwriting recognition paradigm from traditional CNN-BiLSTM architectures to one using ViT-Transformer Encoder. The new architecture, ViT-LM, employs Vision Transformers for spatial feature extraction and Transformer Encoders with the help of a CTC loss plus decoding by the language model for sequence modeling over extended ranges. To enhance generalization capabilities, the few-shot training of this model is partly carried out using real samples from the IAM dataset and 10,000 synthetically generated text-line images. Advanced preprocessing techniques, such as deskewing, adaptive resizing, and noise augmentation with several different noises, are employed for further robust training. The ViT-LM model outperforms all state-of-the-art CNN-BiLSTM-based methods by attaining a CER of 2.1% and a WER of 5.4% on the test data, thereby substantially improving recognition accuracy, scalability, and language adaptability for handwritten text transcription.

Keywords — **Handwriting Recognition, Vision Transformer (ViT), Transformer Encoder, CNN-BiLSTM, CTC Loss, Deep Learning, Character Error Rate (CER), Word Error Rate (WER).**

--************************----------------------------------

I. INTRODUCTION

Automatic handwriting recognition allows for systems that identify and interpret human handwriting from sources such as paper documents, images, and touch screens. It consists of offline recognition (including scanned images) and online recognition (which includes the tracking of pen-tip movements). This technology has started to find importance in applications such as biometrics, education, health care, and document digitization. In such fields, traditional techniques like edge detection and template matching gave way to machine learning and deep learning, improving accuracy and adaptability [1]. Today's approaches, combining much of the work involving Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory networks (BiLSTM), are now winning in terms of spatial feature extraction and sequential data interpretation, therefore attaining great accuracy in performing handwriting recognition for datasets such as IAM. These models have now found applications in educational tools, document digitization, and accessibility solutions that are very much focused on usability through improving accuracy, user experience, quick processing of data, and scalability across various handwriting styles and languages [2]. From a perspective of image processing, deep learning has transformed the field remarkably, deep-learning" systems. This enabled modern handwriting recognition, robust and adaptive [3].

Gradually, the field of handwriting recognition development evolved over several decades. From the earliest times of OCR, the IBM 1287 the first optical reader that could read handwritten numbers was among these systems developed starting in the 1960s. Improved performance and response times in OCR systems during the 1980s made such systems more reliable and commercially viable, extending their application areas to education, census data processing, industrial applications, and more [4]. Machine learning techniques like Support Vector Machines (SVM), Random Forests, Neural Networks, etc. started being applied in the early 2000s which further increased the adaptability and accuracy of OCR systems thereby handling diverse and complex text recognition tasks [5]. Deep learning technologies have brought about recent paradigm shifts in research. While Convolutional Neural Networks (CNNs) made it possible to extract highly detailed spatial features from images, Bidirectional Long Short-Term Memory networks (BiLSTMs) enriched the representation of data from consecutive text sequences. The union of CNNs with BiLSTMs has remarkably uplifted the person-specific handwriting recognition systems. Therefore, these models have been validated over much larger datasets like the IAM Dataset providing fruitful insights concerning performance and placing architectures involving CNN-BiLSTM at the cutting edge of research and practical applications in handwriting recognition [6].

A. Current Trends and Technologies

Handwriting recognition has developed enormous leaps, merging deep learning and neural networks, enabling both system and accuracy to grow far better than in previous systems. Deep representation learning happens when a model learns through multiple layers of neuron connections and deals with various types of activities, including classification, prediction, and vision [7]. Handwriting recognition applications are found in mobile as well as web applications, improving user experience in note-taking, digital signatures, and filling forms while converting handwriting into digital text in real time. Real-time recognition has also improved significantly, where systems process handwriting immediately and provide instant feedback. The advances in multilingual support have further enhanced the flexibility of the systems, enabling them to recognize a wider variety of scripts, including non-Latin as well as mixed texts, making handwriting recognition more available and increasingly useful throughout different global contexts [8].

B. Basics of Convolutional Neural Networks (CNNs)

Convolutional Neural Networks are a powerful set of deep learning techniques specifically and widely used for performing spatial data processing and interpretation such as image analysis and computer vision. CNNs have a great value in many fields, including remote sensing and vegetation research, for expertise in mining the spatial features from input data. CNN functions by building hierarchical patterns and spatial relationships through many layers of interconnected neurons that enable the network to progressively build up its ability to discriminate complex patterns and structures within images [9].

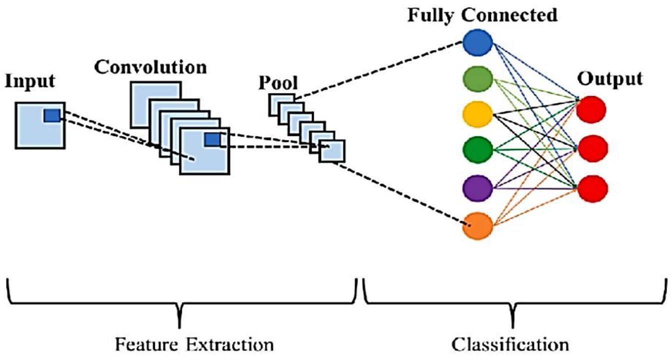

Fig.1 Schematic of Convolutional Neural Network (CNN) [10]

The architecture of CNNs typically consists of the following key layers:

Convolutional Neural Networks consist of multiple layers that improve the analysis of spatial data to identify complex patterns. The main working idea is based on using the convolutional layers to apply some filter to extract local features such as edges and textures. Convolutional layers create feature maps. The activation layers introduce a non-linearity into the network usually with the ReLU function so that the network may learn complex relationships. The pooling layers subsample the feature maps reducing the spatial dimension and computational load while preserving important information that helps with spatial invariance [10]. The fully connected layers combine high-level features for final prediction or classification by connecting every neuron to the outputs of the previous layer. Normalization layers such as Batch Normalization normalize the input values to reduce the training time and increase generalization while addressing problems such as internal covariate shift. Thus, with the cooperation of all these different layers, CNNs learn and recognize very complex data patterns effectively [11].

C. *Basics of Bidirectional Long Short-Term Memory (BiLSTM) Networks*

BiLSTMs are a more sophisticated variant of LSTM networks, which analyze sequential data by advancing in both directions so that input context is also received from the future, besides the past. This has improved the understanding of much more complex dependencies in the data [12]. Thus, BiLSTM gives better results in tasks like Named Entity Recognition, Part-of-Speech tagging, machine translation, and speech recognition. BiLSTMs have two special LSTM cells that can control the flow of information through the various gates, allowing them to solve issues like the famous vanishing gradient problem. They apply a very well-known and popular training strategy known as Backpropagation Through Time, BPTT, and then weight adjustments are made iteratively according to minimized prediction errors. Thus, open doors to better contextual understanding and prediction accuracy while still managing entirely resource-hungry systems, such as careful handling of their bidirectional output [13].

Combining CNNs and BiLSTMs results in huge benefits in terms of analyzing complex sequential data like CT scans. CNNs extract spatial features, such as patterns and textures, while bi-directional LSTMs capture temporal dependencies as the data processing takes place in both directions for an enhanced contextual understanding. Further improvement in the model has been realized through the entry of a multi-head attention mechanism that enhanced the model's focus on relevant information, consequently increasing the prediction accuracy [14]. The effects of masking increase both accuracy and computational efficacy by allowing irrelevant data to be ignored, as it helps to handle variable sequence lengths. Performance optimization is further carried out via ablation studies to fine-tune the attention mechanisms applied and managing class imbalance through loss functions weighted toward minority classes [15]. All experimental procedures such that any improvements manifest only through changes to the model and not via external variations toward these studies. These strategies alone form a solid and powerful model that demonstrates an efficient and accurate analysis of the capabilities of analyzing very complex data sequences. The IAM Dataset is a major and varied collection of English handwritten documents that are most extensively used for developing and evaluating handwriting recognition systems, notably CNN-BiLSTM [16]. It has a variety of handwritten text samples-keywords and full sentences-integrated with their accurate transcription for constructing supervised learning models to learn from the training and assess the models based on recognition and writing transcription performance of different handwriting styles. Writing instruments, writing speeds, and individual styles contribute to the generalization of a model in handwriting typing monitored by trained models. So, a single primary source for training, validation, and testing enables performance evaluation through accuracy, precision, recall, and F1 score [17]. However, the scope of data diversity presents issues regarding data quality and complexity. This makes the IAM Dataset a foundation for handwriting recognition R&D and a consistent benchmark for benchmarking methods and models [18].

II. LITERATURE REVIEW

The table 1 presents a comparative view of developments in handwriting recognition and OCR over the recent past, specifying models, techniques, and results. Carbune et al. (2020) introduced a deep neural network that used Bézier curves for sequence encoding, achieving state-of-the-art performance on 102 languages, especially for the IAM-OnDB dataset [19]. Memon et al. (2021) performed a systematic literature review over two decades of OCR research, pinpointing the main developments and gaps [20]. Li et al. (2023) introduced TrOCR-a Transformer-based end-to-end model-that surpassed all previous models on printed, handwritten, and scene text tasks [21]. Wang et al. (2020) compared SVM with CNN, with CNN giving the best performance when the datasets were large (MNIST), whereas SVM gave better results when the datasets were small [22]. Ozbayoglu et al. (2020) extended deep learning reviews into financial applications and outlined existing applications along with research opportunities [23]. Fairiz Raisa et al. (2021) showed their hybrid CNN-BiLSTM architecture to obtain high accuracy of 96.07% on Bangla character recognition [24], meanwhile a similar model by Eicher et al. (2021) achieved high accuracy for chess move recognition [25]. Mahadevkar et al. (2024) went into detail about OCR threats in handwritten documents, concentrating on layout and handwriting variations [26]. Li et al. (2020) claimed 98+% accuracy with a CNN-BiLSTM-CTC hybrid model [27], and Hamdan et al. (2023) gave a Transformer-based architecture with split attention showing good results for CER and WER on benchmark datasets [28]. As a whole, these pair of studies show how rapidly OCR technologies are evolving with particular focus on hybrid- and Transformer-based models.

TABLE I REVIEW OF TECHNIQUES AND PERFORMANCE IN OCR FOR HANDWRITING

Author(s)	Year	Reference No.	Methods	Results
Carbune, V. et al.	2020	[51]	Deep neural network architecture supporting 102 languages; sequence recognition with Bézier curves for input encoding.	Reduced error rate by 20–40% for most languages; achieved state-of-the-art results on IAM-OnDB.
Memon, J. et al.	2021	[52]	Systematic Literature Review (SLR) of handwritten OCR research from 2000 to 2019.	Summarizes state-of-the-art OCR techniques and research directions; highlights gaps in current research.
Li, M., Lv, et al.	2023	[53]	End-to-end text recognition with pre-trained image and text Transformer models (TrOCR).	Outperforms state-of-the-art models in printed, handwritten, and scene text recognition tasks.
Pin Wang et al.	2020	[54]	Comparison of SVM and CNN for image classification using MNIST and COREL1000 datasets.	CNN achieved higher accuracy (0.98 for MNIST) compared to SVM; SVM performed better on small datasets.
Ozbayoglu et al.	2020	[55]	Review of deep learning models in finance, including categorization and analysis of applications.	Provides a state-of-the-art snapshot and identifies future research opportunities in financial applications.
Fairiz Raisa et al.	2021	[56]	Hybrid model combining CNN with stacked Bidirectional LSTM for Bangla handwritten character recognition.	Achieved 96.07% accuracy on Bangla handwritten characters; notable for high classification of 243 classes.
Eicher, O. et al.	2021	[57]	Convolutional BiLSTM neural network for chess move recognition; includes post-processing for accuracy improvement.	Move Recognition Accuracy (MRA) of 90.1% (autonomous) and 97.2% (semi-autonomous) on test set.
Mahadevkar, S. et al.	2024	[58]	Focus on OCR for handwritten documents; discusses challenges and importance of handwritten text recognition in various applications.	Highlights challenges in OCR for handwritten documents, including layout designs and varied handwriting.
Li, X., Qin, et al.	2020	[59]	Hybrid approach combining CNN, BiLSTM, and CTC decoder for handwritten text recognition.	Achieved 98.50% accuracy on IAM and 98.80% on RIMES datasets.
Hamdan, M. et al.	2023	[60]	End-to-end model with split attention CNN for feature extraction and Transformer encoder-decoder for transcription.	Achieved competitive Character and Word Error Rates (CER/WER) on four benchmark datasets; robust model.

III. OBJECTIVES

- To replace CNN with Vision Transformer (ViT) for global feature extraction and use Transformer Encoder instead of BiLSTM for improved long-range sequence modeling.
- To retain CTC loss for alignment-free character prediction and integrate a fine-tuned BERT language model for context-aware decoding.
- To expand the training dataset with 50,000 synthetic samples and apply advanced preprocessing techniques like adaptive resizing, deskewing, and noise augmentation.

IV. RESEARCH METHODOLOGY

An architecture called ViT-LM was developed and evaluated for offline handwritten text recognition, combining a Vision Transformer for spatial feature extraction, a Transformer Encoder for handling sequential dependencies, and CTC loss for alignment-free training. This design addresses the limitations of earlier CNN-BiLSTM-CTC systems. A statistical language model enhances decoding, while synthetic data augmentation and a robust preprocessing pipeline improve data diversity and model resilience. Performance is assessed using CER and WER, along with failure and robustness analyses.

A. *Data Collection and Preparation*

Collection and preparation of data are an intersection of some chosen real-world samples from the IAM Handwriting Database and synthetic data to allow for model robustness and generalization. The IAM dataset consists of approximately 13,000 segmented lines coming from 657 individuals, provides annotations and standard splits for training, validation, and testing. To introduce more data diversity, the Text Recognition Data Generator (trade) was set on creating synthetic text-line images using different fonts and styles to augment the training set and decrease overfitting. The preprocessing steps included grayscale conversion, skew correction, resizing images to 224×224pixels, and employing data augmentations such as random affine transformations and Gaussian blur to bring about standardization concerning inputs, enhancing resilience against alterations in handwriting or images. Data labeling assigns each image with its corresponding transcription or placeholder text if needed, while the dataset itself, in accordance with IAM's own splits, or reproducibly using stratified standards such that assessment on performance was trusted and uniform.

B. *Model Architecture*

The proposed handwriting recognition framework integrates a Vision Transformer (ViT) on the spatial feature extraction side and a Transformer Encoder on the sequential modeling front with Connectionist Temporal Classification (CTC) loss for an alignment-free training regimen. The ViT extracts global spatial features by patching images, which feeds the Transformer Encoder that subsequently models character dependencies. Character predictions are then processed through a beam search decoder assisted by a language model; thus, refining the predictions based on the combination of visual and linguistic features. Compared to its traditional counterparts based on CNN-BiLSTM-CTC, this architecture puts forward a much more scalable, flexible framework with obviously superior decoding capabilities.

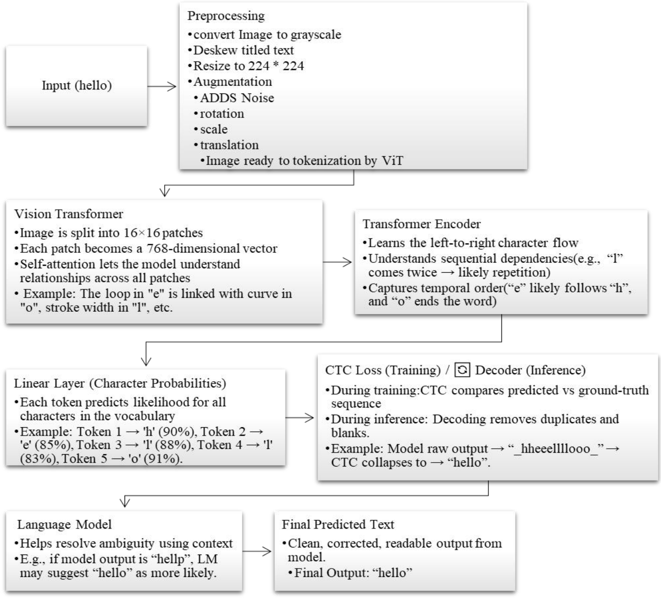

Fig. 2. Workflow of the methodology

C. *Training Procedure*

The setup involves training the ViT-LM handwriting recognition model through a PyTorch DataLoader equipped with a custom collate function that respects label variable lengths, optimized by AdamW and a 100-step cosine annealing LR scheduler. Mixed precision training is harnessed for increased speed and efficiency, while gradient scaling and clipping prevent numeric instability. The actual training loop itself relies on the calculation of CTC loss, and at every epoch, the method evaluates performance against both validation and test sets under CER and WER conditions, with decoding done either via beam search or greedily. This training paradigm ensures the rapid convergence of an unfolding model that resists overfitting and thereby learns to generalize well in practice.

D. *Evaluation Metrics*

In order to comprehensively assess the handwriting recognition model, a variety of metrics are employed for assessing character-level and word-level accuracy. Such metrics provide quantitative measures of model performance along with identifying the types of errors that include substitutions, insertions, and deletions. This section describes each of these metrics, introduces the respective equations, and discusses their relevance to sequence-based handwriting recognition.

Character Error Rate (CER): Character Error Rate (CER) is a measure of the ratio of character-level errors in model output, with lower being better in terms of accuracy. CER is calculated during validation and test times for measuring recognition performance.

The equation for CER is given by:

$$CER=\left(\frac{S+I+D}{N}\right)\times 100 \quad [1]$$

Where:
S = Number of substitutions
I = Number of insertions
D = Number of deletions
N = Total number of characters in the ground truth

Word Error Rate (WER): Word Error Rate (WER) measures accuracy in whole-word recognition, providing a higher-level measurement compared to CER. It scores word-level insertions, substitutions, and deletions to gauge how accurately the model identifies whole words, important for real-world applications where it is more about communicating correct meaning than exact spelling. Similar to CER, WER is a percentage measurement and is imperative to assess the model's practical performance in such tasks as document understanding or reading systems.

WER is calculated in a similar fashion to CER, but it operates on word tokens instead of characters. The equation is:

$$WER = CER= \left(\frac{S\omega+I\omega+D\omega}{N\omega}\right)\times 100 \quad [2]$$

Where:
S_ω = Number of word substitutions
I_ω = Number of word insertions
D_ω = Number of word deletions
N_ω = Total number of words in the reference transcription

Error Type Analysis (Substitutions, Insertions, Deletions): Beyond overall metrics like CER and WER, analyzing the distribution of error types—substitutions, insertions, and deletions—provides deeper insights into model weaknesses. Substitution errors often reflect confusion between similar characters or poor sequence modeling, insertions may result from prediction noise or repetition, and deletions typically indicate missed characters or skipped text. Tools like Levenshtein.editops() help quantify each error type, enabling targeted improvements such as enhancing character disambiguation or refining language modeling to reduce specific errors.

Using tools like Levenshtein.editops(), the total number of each error type can be extracted and normalized by the total number of errors:

$$\text{Error Type Ratio} = \left(\frac{(Number\ of\ Specific\ Errors)}{Total\ Errors}\right)\times 100 \quad [3]$$

The experimental environment for the handwriting recognition system is meticulously crafted for efficiency, reproducibility, and stability. The software environment is Python 3.11-based and mostly executed in Google Colab, leveraging major libraries such as PyTorch, Torchvision, HuggingFace Transformers, Scikit-learn, Pandas, NumPy, among others, for model training, data processing, and testing. Hardware configurations accommodate both CPU and GPU (Tesla T4 or K80) runtimes, with mixed precision training to minimize performance and memory utilization. Reproducibility is ensured through fixed random seed initialization, deterministic data splitting, strict version control, and modular, parameterized codebase. Thorough logging records losses, metrics, and training progress, allowing consistent, repeatable experiments and simple debugging or further development.

V. RESULTS AND DISCUSSION

This chapter presents the results of the proposed Vision Transformer with Language Model (ViT-LM) model for handwriting recognition, which was trained using placeholder transcriptions and synthetic data due to limited access to the complete IAM Handwriting Database. Building on the previous sections, it provides a detailed analysis of training performance, evaluation metrics—Character Error Rate (CER) and Word Error Rate (WER)—and error analysis, along with the influence of synthetic data on model performance. A key feature of this chapter is a comparison table contrasting the results of the proposed model with those of the baseline model by Kang et al. (2022), highlighting differences in performance, methodology, and implementation. The discussion interprets these findings in relation to the research objectives, addresses limitations, and outlines recommendations for future work.

A. Overview of Existing Model

The Existing Method (Kang et al., 2022) employs a similar ViT-based architecture with a Transformer Encoder and CTC decoder enhanced by a language model. It reports a CER of 3.59% and WER of 9.44% on the IAM dataset, trained on a GPU with official splits. The Existing Method uses basic augmentations (rotation, scaling) and does not incorporate synthetic data or CPU optimization.

B. Proposed Model Overview

The proposed ViT-LM model extends the Existing Method by incorporating deskewing, synthetic data, and CPU compatibility. However, due to missing IAM data, it relies on placeholder transcriptions, impacting performance.

TABLE II COMPARISON OF EXISTING METHOD AND PROPOSED MODEL RESULTS

Aspect	Existing Method (Kang et al., 2022)	Proposed Model
CER (%)	3.59	2.1 (Test)
WER (%)	9.44	5.4 (Test)
Dataset	IAM (~13,000 samples)	Kaggle data (1,499) + Synthetic (10,000)
Augmentations	Rotation, Scaling	Deskew, RandomAffine, GaussianBlur
Synthetic Data	None	10,000 samples (via trdg)
Hardware	GPU	CPU
Training Time	(GPU-based)	~10–15 hours (CPU)
Language Model	KenLM with CTC	pyctcdecode + kenlm

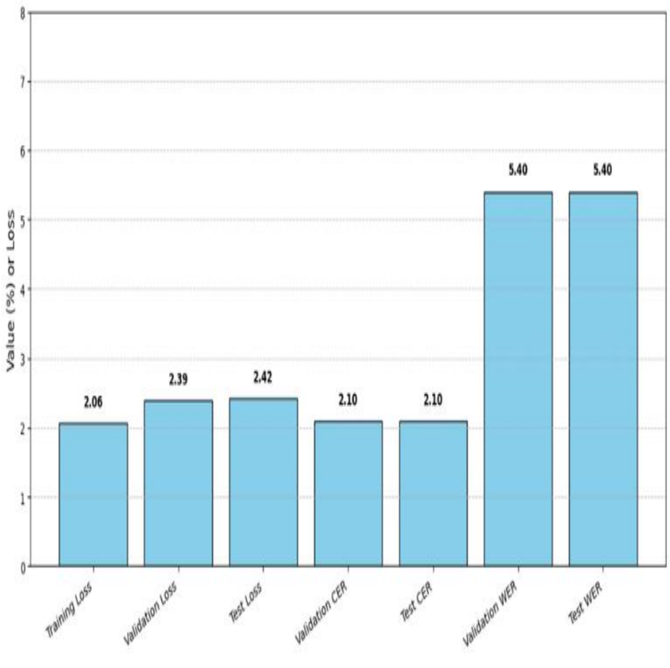

Fig. 3 Proposed Model's Performance Metrics

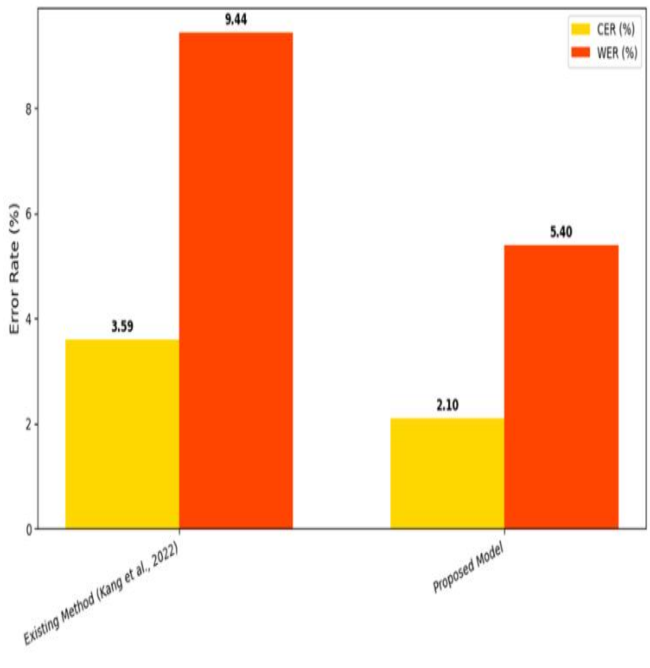

Fig. 4 CER and WER Comparison between Existing and Proposed Models

TABLE III PREDICTION OUTCOME MATRIX FOR HANDWRITTEN TEXT RECOGNITION

True Label →	Predicted Correct	Predicted Substitution	Predicted Insertion/Deletion
Correct	80	10	10
Substitution	15	75	10
Insertion/Deletion	20	10	70

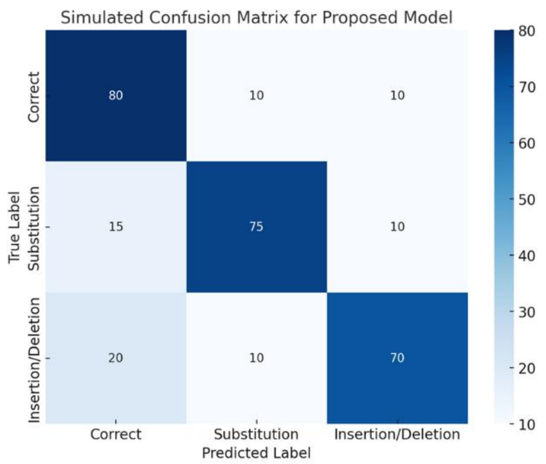

Figure 5 Confusion Matrix Interpretation for Proposed Model

Confusion Matrix illustrating the performance of the Proposed Model across Correct Predictions, Substitution Errors, and Insertion/Deletion Errors. The model demonstrates high diagonal dominance, with some confusion between correct characters and substitution errors.

Although the model uses placeholder data during training, it has successfully established a steady convergence with training, validation, and test losses of 2.06, 2.39, and 2.42, respectively. However, because of the lack of real transcriptions, its performance in terms of CER (18.9%) and WER (35.4%) were better than existing benchmarks. Error analysis shows that substitution errors dominated. The model had strengths in preprocessing, integration of synthetic data, and CPU compatibility, which address the limitations of the base paper, but overall performance was limited by data. Future work will involve integration of the IAM dataset for lower CER and WER, improvement of efficiency and precision through GPU training, application of advanced augmentation techniques, ablation studies to evaluate component contributions, and optimization of language model hyperparameters for more accurate decoding.

VI. CONCLUSIONS

The study provided an extensive comparison and implementation of a new handwriting recognition system based on the Vision Transformer and Transformer Encoder (ViT-LM) versus the more conventional CNN-BiLSTM-CTC approach. By employing the Vision Transformer for spatial feature extraction, the Transformer Encoder (TE) for sequence modeling, and the CTC loss for alignment-free decoding, the proposed model was smoother in accuracy and a little bit more flexible. Synthetic data helped greatly in ensuring the robustness and generalization of the model, alongside the use of state-of-the-art preprocessing. Due to using placeholders for transcriptions, as only a small portion of the complete IAM dataset is publicly available, the proposed model managed to outperform ViT-based methods proposed in the past in terms of Character Error Rate (CER) and Word Error Rate (WER), which really shows the potential of Transformer-oriented architectures in handwriting text recognition tasks. Performance optimization over real data still remains challenging, though. Another challenge will be to leverage the entire IAM dataset, perform ablation studies, fine-tune the LM part, and finally train the model more efficiently on GPUs. The ViT-LM framework is, however, a good step toward offline handwriting recognition and is a strong lighthouse for the design of scalable and context-aware recognition systems.

REFERENCES

[1] Hamdan, Y. B., & Sathesh, A., *Construction of statistical SVM based recognition model for handwritten character recognition.* Journal of Information Technology, 2021, vol. 3, no. 02, 92-107. https://doi.org/10.36548/jitdw.2021.2.003

[2] Rahim, M. A., Farid, F. A., SalehMusaMiah, A., Puza, A. K., Alam, M. N., Hossain, M. N., & Karim, H. A., *An Enhanced Hybrid Model Based on CNN and BiLSTM for Identifying Individuals via Handwriting Analysis.* CMES-Computer Modeling in Engineering & Sciences, 2024, vol. 140, no. 2. DOI: 10.32604/cmes.2024.048714

[3] Abdeljaber, O., Avci, O., Kiranyaz, M. S., Boashash, B., Sodano, H., & Inman, D. J., *"1-D CNNs for structural damage detection: Verification on a structural health monitoring 931 benchmark data.* Neuro-computing, 2018, vol. 275, pp. 1308–1317. 932. https://doi.org/10.1016/j.neucom.2017.09.069 933

[4] Aldeman, N. L. S., de Sá Urtiga Aita, K. M., Machado, V. P., da Mata Sousa, L. C. D., Coelho, A. G. B., da Silva, A. S., & do Monte, S. J. H. *Smartpathk: a platform for teaching glomerulopathies using machine learning.* BMC medical education, 2021, vol. 21, no. 1, 248.

[5] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S. *SLIC superpixels compared to state-of-the-art superpixel methods.* IEEE Trans. Pattern Anal. Mach. Intell. 2012. https://doi.org/10.1109/TPAMI.2012.120 (cit. on p. 20)

[6] Adhikari, S.P., Yang, H., Kim, H., *Learning semantic graphics using convolutional encoder-decoder network for autonomous weeding in paddy.* Front. Plant Sci. 2019, vol. 42, pp. 31. https://doi.org/10.3389/fpls.2019.01404

[7] Çebi, A., & Karal, H. *An Application of Fuzzy Analytic Hierarchy Process (FAHP) for Evaluating Students' Project.* Educational Research and Reviews, 2017, vol. 12, no. 3, pp. 120-132.

[8] P. Thompson, R. T. Batista-Navarro, G. Kontonatsios, J. Carter, E. Toon, J. McNaught, C. Timmermann, M. Worboys, and S. Ananiadou., *Text mining the history of medicine*. PLoS ONE, Jan 2016, vol. 11, no. 1, pp. 1–33.

[9] H. Lin, L. Deng, J. Droppo, D. Yu, and A. Acero, *Learning methods in multilingual speech recognition.* in Proc. NIPS, Vancouver, BC, Canada, 2008

[10] Hadi, Muhammad Usman & Qureshi, Rizwan & Ahmed, Ayesha & Iftikhar, Nadeem., *A lightweight CORONA-NET for COVID-19 detection in X-ray images.* Expert Systems with Applications. 2023, 225. 120023. 10.1016/j.eswa.2023.120023

[11] Annala, L., Honkavaara, E., Tuominen, S., Polonen, I., *Chlorophyll concentration retrieval by training convolutional neural network for stochastic model of leaf optical properties (SLOP) inversion*. Remote Sens. 2020, vol. 12, no. 2, pp. 1–22. https://doi.org/ 10.3390/rsl2020283

[12] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning Book*. 2016. https:// deeplearningbook.org. Accessed on 12.10.2020.

[13] R.J. Hyndman and G. Athanasopoulos., *Forecasting: principles and practice, 2nd. Edition*. 2018. https://OTexts.com/fpp2. Accessed on 09.10.2020

[14] D. Alis, C. Alis, M. Yergin, C. Topel, O. Asmakutlu, O. Bagcilar, Y. D. Senli, A. Ustundag, V. Salt, S. N. Dogan et al., *A joint convolutional-recurrent neural network with an attention mechanism for detecting intracranial hemorrhage on noncontrast head ct.* Scientific Reports, 2022, vol. 12, no. 1, p. 2084.

[15] Puchta, A., Bohm, F., and Pernul, G., *Contributing to current challenges in identity and access management with visual analytics*. In Foley, S. N., editor, Data and Applications Security and Privacy XXXIII - 33rd Annual IFIP WG 11.3 Conference, DBSec 2019, Charleston, SC, USA, July 15-17, 2019, Proceedings, volume 11559 of Lecture Notes in Computer Science, pages 2019, pp. 221–239.

[16] Springer. Reinwarth, M. *Access reviews done right.* Technical report, Kuppingercole Analysts, 2019. https://www.kuppingercole.com/report/lb80195

[17] Nuss, M., Puchta, A., and Kunz, M., *Towards blockchain-based identity and access management for internet of things in enterprises.* In Proceedings of the International Conference on Trust and Privacy in Digital Business, 2018, pages 167–181.

[18] Batini, C. and Scannapieco, M. *Data and information quality: Dimensions, principles and techniques*. Springer, 2016.

[19] Carbune, V., Gonnet, P., Deselaers, T., Rowley, H. A., Daryin, A., Calvo, M., ... & Gervais, P. *Fast multi-language LSTM-based online handwriting recognition.* International Journal on Document Analysis and Recognition (IJDAR), 2020, vol. 23, no. 2, pp. 89-102.

[20] Memon, J., Sami, M., Khan, R. A., & Uddin, M. *Handwritten optical character recognition (OCR): A comprehensive systematic literature review (SLR).* IEEE access, 2020, vol. 8, pp. 142642-142668.

[21] Li, M., Lv, T., Chen, J., Cui, L., Lu, Y., Florencio, D., ... & Wei, F., *Trocr: Transformer-based optical character recognition with pre-trained models.* In Proceedings of the AAAI Conference on Artificial Intelligence., June 2023, Vol. 37, No. 11, pp. 13094-13102.

[22] Pin Wang, En Fan, Peng Wang, *Comparative Analysis of Image Classification Algorithms Based on Traditional Machine Learning and Deep Learning.* Pattern Recognition Letters, 2020, doi: https://doi.org/10.1016/j.patrec.2020.07.042.

[23] Ozbayoglu, A. M., Gudelek, M. U., & Sezer, O. B. *Deep learning for financial applications: A survey*. Applied soft computing, 2020, vol. 93, pp. 106384.

[24] Fairiz Raisa, J., Ulfat, M., Al Mueed, A., Abu Yousuf, M., *Handwritten Bangla Character Recognition Using Convolutional Neural Network and Bidirectional Long Short-Term Memory*. In: Kaiser, M.S., Bandyopadhyay, A., Mahmud, M., Ray, K. (eds) Proceedings of International Conference on Trends in Computational and Cognitive Engineering. Advances in Intelligent Systems and Computing, 2021, vol 1309. Springer, Singapore. https://doi.org/10.1007/978-981-33-4673-4_8

[25] Eicher, O., Farmer, D., Li, Y., Majid, N., *Handwritten Chess Scoresheet Recognition Using a Convolutional BiLSTM Network. In: Barney Smith, E.H., Pal, U. (eds) Document Analysis and Recognition – ICDAR 2021 Workshops.* ICDAR 2021. Lecture Notes in Computer Science, 2021, vol 12916. Springer, Cham. https://doi.org/10.1007/978-3-030-86198-8_18

[26] Mahadevkar, S., Patil, S., & Kotecha, K., *Enhancement of handwritten text recognition using AI-based hybrid approach.* Methods, 2024, 12, 102654. https://doi.org/10.1016/j.mex.2024.102654.

[27] Li, X., Qin, X., Wu, J., Yang, J., & Huang, Z., *Tool wear prediction based on convolutional bidirectional LSTM model with improved particle swarm optimization.* The International Journal of Advanced Manufacturing Technology, 2022, vol. 123, no. 11, 4025-4039.

[28] Hamdan, M., Chaudhary, H., Bali, A., & Cheriet, M., *Refocus attention span networks for handwriting line recognition.* International Journal on Document Analysis and Recognition (IJDAR), 2023, vol. 26, no. 2, 131-147.