IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 3, June 2026, pp. 2655~2663
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i3.pp2655-2663

Vision transformer and hybrid models for Malayalam handwritten word recognition

Anju Arangil Thazhath[1,2], Binu Poothakuzhiyil Chacko[3], Mohamed Basheer Kizhakke Parambath[4]
[1]Department of Computer Science, Sullamussalam Science College, University of Calicut, Kerala, India
[2]Department of Computer Science, Faculty of Data Science and Analytics, University of Calicut, Kerala, India
[3]Department of Computer Science, Prajyoti Niketan College, Kerala, India
[4]Department of Computer Science, Amal College of Advanced Studies, Kerala, India

Article Info

Article history:

Received Oct 12, 2024
Revised Apr 1, 2026
Accepted Apr 21, 2026

Keywords:

Attention mechanism
Feed-forward neural network
Handwritten word recognition
Malayalam word dataset
Vision transformer

ABSTRACT

Transformer-based architectures and attention mechanisms have revolutionized the field of image recognition. This study focuses on offline handwritten Malayalam word recognition, addressing the lack of publicly available datasets for this low-resource language. A new Malayalam word dataset (MWD) comprising 20,850 samples across 139 classes was developed to support research in this domain. The vision transformer (ViT) was employed for advanced feature extraction, and multiple recognition models—feed-forward neural network (FFNN), global average pooling (GAP), bidirectional long short-term memory (BiLSTM), and attention-based feed-forward neural network (AFFNN)—were evaluated. Among these, AFFNN achieved the highest accuracy of 98.56%, establishing the proposed vision transformer-based attention handwritten word recognition (ViTA-HWR) model as a robust framework for handwritten Malayalam word recognition and valuable contribution to regional language processing.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Anju Arangil Thazhath
Department of Computer Science, Sullamussalam Science College, University of Calicut
Areekode, Malappuram, Kerala, India
Email: anjuat@sscollege.ac.in

1. INTRODUCTION

Optical character recognition (OCR) has long served as a vital technology for converting printed or machine-generated text into digital formats. However, as the need for digitization expanded, attention shifted toward offline handwritten word recognition (HWR), a challenging subset of OCR that deals with handwritten text. Unlike printed characters, handwritten words exhibit substantial variability due to differences in individual writing styles, pressure, stroke width, and character connectivity. This complexity intensifies for scripts such as Malayalam, a Dravidian language spoken predominantly in the Indian state of Kerala, known for its compound characters, rounded shapes, and intricate ligatures. HWR plays a crucial role in applications such as document digitization, information retrieval from handwritten records, and preservation of historical manuscripts, yet its effectiveness is often hindered by script-specific challenges and the absence of robust annotated datasets.

Despite remarkable progress in OCR and handwritten text recognition across languages such as English, Arabic, and Chinese, Malayalam remains a low-resource language, with limited publicly available datasets for handwritten words. The absence of such datasets impedes the development and benchmarking of reliable recognition models. Furthermore, Malayalam handwritten words often vary in spacing and curvature, leading to segmentation difficulties and recognition errors. Therefore, developing a specialized dataset and an efficient recognition framework is critical for advancing research in Malayalam handwritten

text processing [1]. Earlier works in HWR primarily relied on handcrafted feature extraction techniques, where geometric, statistical, and texture-based features were used in combination with traditional machine learning classifiers [2]–[7]. The advent of deep learning introduced powerful architectures such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) networks, which significantly improved recognition accuracy by automatically learning spatial and sequential dependencies [8]–[18]. More recently, transformer-based architectures have revolutionized the field by capturing long-range dependencies and enabling parallelized feature computation. Studies such as transformer-based optical character recognition (TrOCR) for English handwritten text [19] and optical character former (OCFormer) for Arabic handwriting have demonstrated the superior performance of transformer-based models in recognizing diverse handwriting styles [18], [20], [21]. In the context of Malayalam, existing studies remain limited to CNN-based feature extraction coupled with conventional classifiers such as support vector machine (SVM) or feed-forward neural networks (FFNN) [15], highlighting a significant gap for advanced transformer-based approaches.

In response to these limitations, the present study introduces a comprehensive Malayalam word dataset (MWD) comprising 20,850 handwritten word samples across 139 classes, each containing 150 balanced samples. Furthermore, a novel recognition framework termed vision transformer-based attention handwritten word recognition (ViTA-HWR) is proposed. The model integrates a vision transformer (ViT) for feature extraction and an attention-based feedforward neural network (AFFNN) for recognition. The ViT module captures holistic global representations of handwritten word images, while the attention mechanism in AFFNN dynamically emphasizes the most discriminative features, leading to robust recognition even in complex handwritten scripts.

The key contributions of this research are fourfold: i) development of a dedicated Malayalam handwritten word dataset, which serves as a new benchmark for the research community; ii) introduction of a transformer-based hybrid recognition architecture combining ViT and attention-enhanced FFNN for improved performance; iii) comprehensive evaluation of multiple recognition strategies, including FFNN, global average pooling (GAP), bidirectional long short-term memory (BiLSTM), and AFFNN, revealing the superiority of attention mechanism; and iv) achievement of 98.56% accuracy on proposed dataset, marking a significant advancement in Malayalam HWR. This study thus bridges the research gap for low-resource scripts and contributes both a new dataset and an innovative transformer-driven recognition framework. Overall, it sets a foundation for future advancements in regional handwriting recognition systems.

2. METHOD

This section describes the systematic workflow followed in this study. The workflow includes dataset creation, feature extraction using the ViT, and HWR through various deep learning models. The overall process is illustrated in Figure 1.

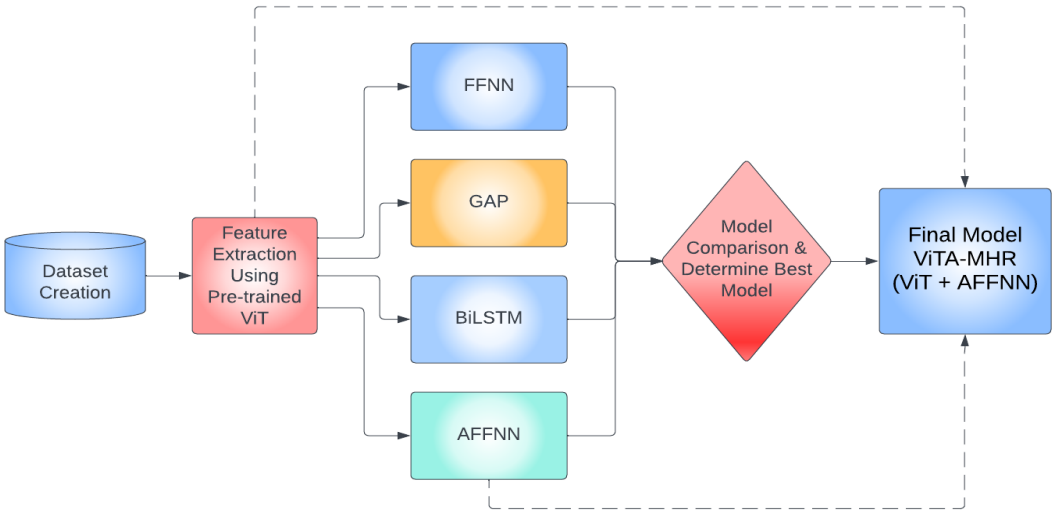

Figure 1. Workflow diagram of the proposed HWR system

2.1. Dataset creation

Since no publicly available handwritten MWD exists, a new dataset titled MWD was created to facilitate training and evaluation of the proposed models. The dataset was constructed using handwritten police first information statements (FIS), which were chosen due to their practical importance and the consistent handwritten format used across Kerala Police Departments. The FIS documents were sourced from the District Crime Record Bureau (DCRB) in Kozhikode and scanned into digital format for processing. To extract individual word samples, an image processing pipeline was applied, involving contour detection and bounding box rectangle methods. In this work, the dilation process was modified to connect individual words rather than entire text lines, enhancing segmentation precision. Subsequently, word clustering techniques were used to group segmented words into distinct clusters representing unique word classes.

The final MWD comprises 20,850 handwritten word samples, distributed evenly across 139 classes, with each class containing 150 samples. Word selection was guided by the linguistic frequency and contextual significance of words in Malayalam documents, ensuring broad representational coverage. To ensure style diversity, handwriting samples were collected from 750 individuals aged 10–65 years, representing different educational and regional backgrounds. This diversity improves model generalization by exposing it to a wide range of handwriting variations. Quality control checks were performed to eliminate incomplete or noisy samples. Figure 2 shows representative samples from the developed MWD dataset.

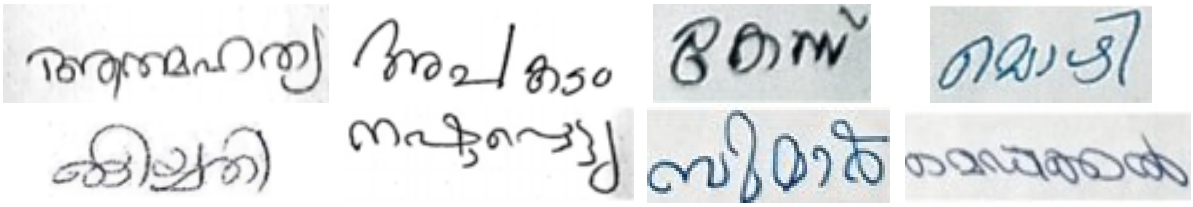

Figure 2. Sample words from MWD dataset

2.2. Research workflow

Recognizing handwritten Malayalam words is a complex problem due to the compound script structure, writing variations, and cursive nature of the language. To address these challenges, a transformer-based framework was developed that combines the ViT for feature extraction with different recognition models. The experimental workflow is summarized in Algorithm 1. Empirical evaluation revealed that the AFFNN outperformed the other models, leading to the final hybrid architecture termed ViTA-HWR.

Algorithm 1. Malayalam HWR workflow
Input: Handwritten word dataset with 139 classes, each containing 150 samples.
Output: Recognized Malayalam handwritten words.

1: Dataset preparation: perform preprocessing and segmentation to obtain individual word samples from scanned handwritten FIS documents.
2: Feature extraction using ViT: transform images into patches and extract features using ViT.
Form a feature matrix $X \in R^{N \times F}$, where N is the number of samples and F the feature dimension.
3: Recognition models:

- FFNN

$$O_{FFNN} = \sigma\,(W_{FFNN}X + b_{FFNN})$$

- GAP

$$O_{GAP} = GlobalAvgPooling(X)$$

- BiLSTM

$$O_{BLSTM} = \sigma\,(W_{LSTM}X + b_{BLSTM})$$

- AFFNN

$$O_{ATT} = \sigma\,(W_{ATT}(X \odot A) + b_{ATT})$$

Where A denotes the attention matrix, $\odot$ element-wise multiplication, and σ the activation function.

4: Comparative evaluation: assess all models using metrics such as accuracy, precision, recall, and F1-score.
5: Model selection: identify the model with the highest recognition performance.

2.3. Feature extraction using vision transformer

The ViT was equipped to extract discriminative visual features from handwritten word images. Unlike CNNs, which focus on local patterns via convolutional kernels, ViT captures global contextual dependencies through the self-attention mechanism. This thereby improves feature representation and recognition robustness [22].

2.3.1. Vision transformer architecture

The ViT processes an image as a sequence of fixed-size patches [23]. Each input image X is divided into patches as $x_{i,j} = P(X)$, where $P(\cdot)$ denotes the patch extraction operation. Positional embeddings are then added to preserve spatial information: $x_{i,j} = x_{i,j} + PE(i,j)$, where $PE(i,j)$ is the positional encoding for the patch located at (i,j). The resulting token sequence is linearly projected and passed through multiple transformer encoder layers, each comprising multi-head self-attention and feedforward sub-layers. For each attention head, the scaled dot-product attention is computed as in (1).

$$Attention(Q,K,V) = softmax\left(QK^T\middle|\sqrt{d_k}\right)V \quad (1)$$

Where Q, K, and V represent the query, key, and value matrices, and d_k is the key dimension. The multi-head mechanism allows the model to attend to different subspaces of the feature space, enhancing its ability to model relationships between distant image regions. The output feature embeddings are normalized and used as input to the recognition module. Figure 3 depicts the ViT architecture.

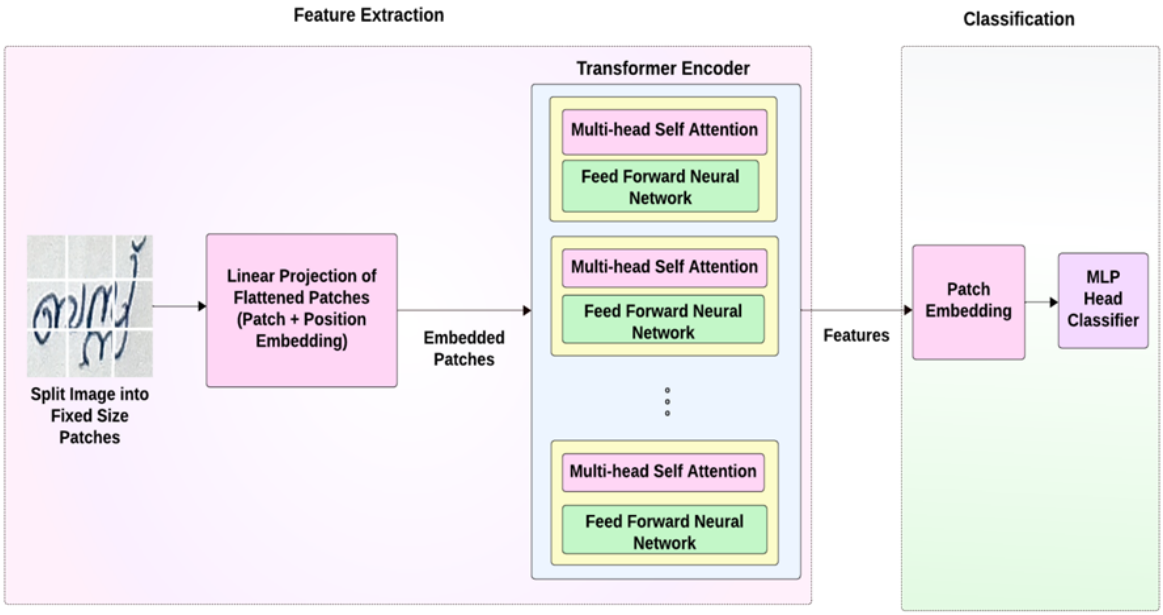

Figure 3. ViT architecture

2.4. Recognition models

Following ViT-based feature extraction, four distinct recognition strategies were implemented to evaluate performance differences.

i) FFNN: the FFNN comprises two fully connected layers. The first layer reduces feature dimensionality:$h1 = f(W1 \times x + b1)$, where $W1$ is the weight matrix of the first layer, $b1$ is the bias vector for the first layer, and f is the activation function. The second layer projects to class scores: $y = W2 \times h1 + b2$, where $W2$ is the weight matrix of the second layer, $b2$ is the bias vector for the second layer, y is the output vector, where each element corresponds to a class score [24].

ii) GAP: the GAP aggregates each feature channel by averaging: $v_c = \frac{1}{N}\sum_{i=1}^{N} x_{i,c}$. Then, a dense layer with softmax activation classifies the pooled representation [25].

iii) BiLSTM: the BiLSTM to exploit sequential dependencies within features, BiLSTM processes input in both directions: $h_t = BLSTM(x_t, h_{t-1})$. The final hidden state is passed through dense layers for classification [25].

iv) AFFNN: this model embeds multi-head self-attention mechanism within FFNN to emphasize crucial spatial features. For each attention head: $Head_i = Attention\ (Q_i, K_i, V_i)$. The outputs from all heads are concatenated and linearly transformed: $Final\ Attention\ Output = LinearTransform\ (Concat(Head_1, Head_2, Head_3, Head_4))$. A residual connection ensures information preservation: $Output = X + SelfAttention(X)$. Two dense layers then perform classification. The attention mechanism enables the network to dynamically focus on the most informative features, improving robustness against handwriting variations and noise.

2.5. Proposed vision transformer-based attention handwritten word recognition model

The final architecture, named ViTA-HWR, integrates ViT-based feature extraction with the AFFNN recognition module as shown in Figure 4. This hybrid approach captures both global contextual patterns and localized discriminative details. It ensures efficient and accurate recognition of Malayalam handwritten words.

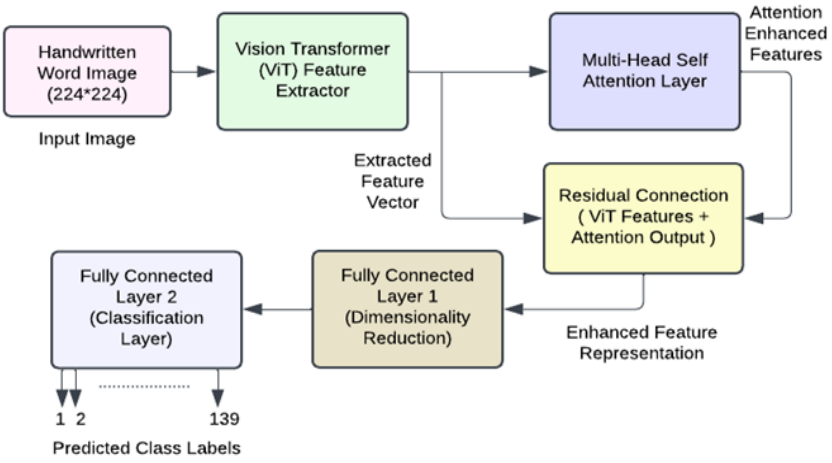

Figure 4. ViTA-HWR architecture

3. RESULTS AND DISCUSSION

This study comprehensively compares four recognition strategies: FFNN, average pooling, BiLSTM, and AFFN. Notably, the attention-based approach stands out with remarkable performance. This highlights that integrating an attention mechanism improves the model's capacity to identify intricate patterns within Malayalam handwritten words. Through comparative experiments, the ViTA-HWR achieved the highest recognition accuracy of 98.56%, demonstrating the effectiveness of combining transformer-based feature extraction with attention-driven recognition. Table 1 summarizes the experimental results of all tested models.

Table 1. Experimental results

Model	Validation loss	Validation accuracy (%)	Test loss	Test accuracy (%)
AFFNN	0.2432	98.66	0.2364	98.56
FFNN	0.1068	97.54	0.0957	97.89
BLSTM	0.1444	97.03	0.1483	97.09
GAP	0.2031	95.14	0.1950	95.53

3.1. Comparison with state-of-the-art techniques

The comparative analysis presented in the above table highlights the evolution of HWR approaches across Bengali and Malayalam scripts. These scripts are known for their complex character structures and similar stroke patterns, making them suitable for cross-script comparison. Earlier methods primarily relied on traditional classifiers like SVM with hand-crafted features. At the same time recent studies have shifted

towards deep learning-based approaches such as CNN and FFNN, coupled with data augmentation strategies to improve generalization. Despite these advancements, most existing works either focus on limited datasets or do not incorporate attention mechanisms for enhanced feature discrimination.

In contrast, the proposed ViTA-HWR model introduces a hybrid architecture that combines the global feature extraction capability of ViT with the adaptive learning strength of attention-based feedforward networks. This integrated design eliminates the need for data augmentation and achieves superior accuracy, thereby demonstrating a significant advancement over existing models. Table 2 compares the presented work with existing offline HWR works in both Malayalam and Bengali languages.

Table 2. Comparison of state-of-the-art offline HWR studies and proposed work

Source	Dataset (samples)	Data augmentation	Feature extraction	Recognition	Accuracy (%)
Bhowmik *et al.* [6]	18,000	Nil	Shape-based feature descriptor	SVM	83.64
Malakar *et al.* [17]	18,000	Nil	CNN	CNN model, lottery Ticket hypothesis	91.30
Das *et al.* [16]	18,000	Nil	CNN	FFNN (HWordNet)	96.17
Jino *et al.* [15]	29,516	Rotation, Gaussian noise, shifting	CNN	SVM	96.90
Proposed work	20,850	Nil	ViT	AFFNN	98.56

To establish the effectiveness of the proposed ViTA-HWR model, a comparative analysis was conducted against existing benchmark models for HWR. Two state-of-the-art approaches were selected based on their relevance to the task:

i) Pre-trained CNN + SVM model: this model was chosen as it represents the state-of-the-art approach for Malayalam HWR, utilizing a pre-trained CNN for feature extraction followed by an SVM classifier for recognition [15]. It provides a direct comparison with existing methods designed specifically for the Malayalam script. This model achieved a test accuracy of 87.85% on the MWD.

ii) HWordNet CNN model: a well-established benchmark model for low-resource languages, particularly Bangla, which shares structural similarities with Malayalam regarding complex script-based writing styles [16]. Since Bangla and Malayalam both present challenges in handwritten text recognition, evaluating HWordNet on the MWD allows for a broader assessment of its applicability. This model achieved a test accuracy of 96.69%.

The results indicate that the proposed ViTA-HWR model significantly outperforms both benchmark models, demonstrating superior recognition accuracy. The ViTA-HWR model's attention-enhanced feature extraction and classification framework effectively capture the complexities of handwritten Malayalam words, further reinforcing its suitability for low-resource handwritten text recognition. To ensure reproducibility, proposed model was trained with learning rate of 0.0001, batch size of 32, and for 10 epochs.

3.2. Benchmarking the model on CMATERdb2.1.2 public dataset

To evaluate the generalization capability of the proposed model, an additional experiment was conducted using the publicly available CMATERdb2.1.2 dataset [5]. This dataset consists of handwritten Bangla word samples representing 120 city names from West Bengal, India. The data was collected in an offline setting and includes diverse handwriting styles influenced by factors such as writer demographics, including age, gender, educational background, and profession. These inherent variations make CMATERdb2.1.2 a challenging benchmark for HWR.

The model was trained and tested on CMATERdb2.1.2 and the results demonstrate that the model effectively adapts to this new dataset, achieving a test accuracy of 97.71%. Additionally, performance metrics further validate the model's reliability, with a micro precision of 0.9811, micro recall of 0.9811, and micro F1-score of 0.9811. The macro precision, macro recall, and macro F1-score were recorded as 0.9818, 0.9815, and 0.9808, respectively.

3.3. Cross-validation analysis

To evaluate the robustness and generalizability of the proposed recognition model, a five-fold cross-validation experiment was conducted on both the private Malayalam handwritten word dataset and the public CMATERdb2.1.2 Bangla dataset. This approach helped minimize data variability effects and provided a reliable assessment of model performance across different training-validation splits. The dataset was

randomly partitioned into five folds, where each fold served as a validation set once, while the remaining four were used for training. This process was repeated for all five folds, and the performance metrics were averaged along with their standard deviation to obtain a more reliable estimate of model accuracy and loss. The training procedure incorporated an early stopping mechanism to prevent overfitting. Table 3 shows the five-fold cross-validation results for both datasets.

Table 3. Five-fold cross-validation results on MWD and CMATERDB2.1.2

Fold	MWD			CMATERdb2.1.2 Bangla dataset		
	Validation accuracy (%)	Validation loss	Epochs	Validation accuracy (%)	Validation loss	Epochs
Fold 1	97.82	0.0858	9	97.76	0.0934	9
Fold 2	99.14	0.0354	6	99.45	0.0163	8
Fold 3	99.14	0.0360	6	99.70	0.0127	6
Fold 4	98.83	0.0925	6	99.42	0.0232	8
Fold 5	99.57	0.0158	5	99.86	0.0087	4
Average	98.90	0.0531		99.24	0.0309	
Standard deviation	0.59	0.0304		0.76	0.0316	

The five-fold cross-validation results demonstrate the proposed recognition model's high reliability and generalization capability across both datasets. The minimal standard deviation in accuracy and loss further confirms the model's stability across different folds. These findings indicate that the model effectively learns discriminative features, ensuring robust recognition across diverse handwritten scripts.

4. CONCLUSION

This study presents a significant advancement in offline HWR by integrating ViT capabilities for feature extraction. The proposed ViTA-HWR model effectively identifies and extracts essential features, enabling a more comprehensive evaluation through four recognition strategies. The comparative analysis demonstrated that the AFFNN model achieved superior performance, attributed to its attention mechanism that dynamically assigns importance to different regions within the feature matrix. This approach enhances the model's ability to capture crucial patterns in handwritten text. Future work will focus on further refining the proposed attention mechanism and extending its evaluation across more diverse and challenging datasets to enhance robustness and adaptability. In addition, the comparative analysis will be expanded to include recent transformer-based and hybrid architectures, such as TrOCR and swin transformer, enabling a more comprehensive assessment of performance trade-offs. A deeper investigation into computational efficiency, error characteristics, and robustness under challenging conditions will also be conducted to better understand the model's practical applicability in real-world scenarios.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Anju Arangil Thazhath	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
Binu Poothakuzhiyil Chacko	✓			✓		✓				✓		✓	✓	
Mohamed Basheer Kizhakke Parambath				✓	✓		✓			✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [AAT], upon reasonable request.

REFERENCES

[1] K. Manjusha, M. A. Kumar, and K. P. Soman, "On developing handwritten character image database for Malayalam language script," *Engineering Science and Technology, an International Journal*, vol. 22, no. 2, pp. 637–645, 2019, doi: 10.1016/j.jestch.2018.10.011.

[2] M. Cheriet, N. Kharma, C.-L. Liu, and C. Suen, *Character recognition systems: A guide for students and practitioners*. John Wiley & Sons, 2007.

[3] S. Madhvanath and V. Govindaraju, "The role of holistic paradigms in handwritten word recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 149–164, 2001, doi: 10.1109/34.908966.

[4] S. Malakar, S. Sahoo, A. Chakraborty, R. Sarkar, and M. Nasipuri, "Handwritten Arabic and Roman word recognition using holistic approach," *Visual Computer*, vol. 39, no. 7, pp. 2909–2932, Jul. 2023, doi: 10.1007/s00371-022-02500-7.

[5] S. Malakar, M. Ghosh, S. Bhowmik, R. Sarkar, and M. Nasipuri, "A GA based hierarchical feature selection approach for handwritten word recognition," *Neural Computing and Applications*, vol. 32, no. 7, pp. 2533–2552, Apr. 2020, doi: 10.1007/s00521-018-3937-8.

[6] S. Bhowmik, S. Malakar, R. Sarkar, S. Basu, M. Kundu, and M. Nasipuri, "Off-line Bangla handwritten word recognition: a holistic approach," *Neural Computing and Applications*, vol. 31, no. 10, pp. 5783–5798, Oct. 2019, doi: 10.1007/s00521-018-3389-1.

[7] S. Sahoo *et al.*, "Handwritten Bangla word recognition using negative refraction based shape transformation," *Journal of Intelligent and Fuzzy Systems*, vol. 35, no. 2, pp. 1765–1777, 2018, doi: 10.3233/JIFS-169712.

[8] K. Sharma, P. K. Sarangi, L. Rani, G. Singh, A. K. Sahoo, and B. P. Rath, "Handwritten digit classification using HOG features and SVM classifier," in *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2022*, 2022, pp. 2071–2074. doi: 10.1109/ICACITE53722.2022.9823782.

[9] N. Alzrrog, J. F. Bousquet, and I. El-Feghi, "Deep learning application for handwritten Arabic word recognition," in *Canadian Conference on Electrical and Computer Engineering*, 2022, pp. 95–100, doi: 10.1109/CCECE49351.2022.9918375.

[10] S. Khosravi and A. Chalechale, "Chimp optimization algorithm to optimize a convolutional neural network for recognizing Persian/Arabic handwritten words," *Mathematical Problems in Engineering*, vol. 2022, pp. 1–12, Apr. 2022, doi: 10.1155/2022/4894922.

[11] A. Zohrevand and Z. Imani, "Holistic Persian handwritten word recognition using convolutional neural network," *International Journal of Engineering, Transactions B: Applications*, vol. 34, no. 8, pp. 2028–2037, 2021, doi: 10.5829/ije.2021.34.08b.24.

[12] F. Abdurahman, E. Sisay, and K. A. Fante, "AHWR-Net: offline handwritten Amharic word recognition using convolutional recurrent neural network," *SN Applied Sciences*, vol. 3, no. 8, pp. 1–11, Aug. 2021, doi: 10.1007/s42452-021-04742-x.

[13] R. Pramanik and S. Bag, "Handwritten Bangla city name word recognition using CNN-based transfer learning and FCN," *Neural Computing and Applications*, vol. 33, no. 15, pp. 9329–9341, Aug. 2021, doi: 10.1007/s00521-021-05693-5.

[14] M. Bonyani, S. Jahangard, and M. Daneshmand, "Persian handwritten digit, character and word recognition using deep learning," *International Journal on Document Analysis and Recognition*, vol. 24, no. 1–2, pp. 133–143, 2021, doi: 10.1007/s10032-021-00368-2.

[15] P. J. Jino, K. Balakrishnan, and U. Bhattacharya, "Offline handwritten Malayalam word recognition using a deep architecture," in *Advances in Intelligent Systems and Computing*, Springer Singapore, 2019, pp. 913–925. doi: 10.1007/978-981-13-1592-3_73.

[16] D. Das, D. R. Nayak, R. Dash, B. Majhi, and Y. D. Zhang, "H-WordNet: a holistic convolutional neural network approach for handwritten word recognition," *IET Image Processing*, vol. 14, no. 9, pp. 1794–1805, Jul. 2020, doi: 10.1049/iet-ipr.2019.1398.

[17] S. Malakar, S. Paul, S. Kundu, S. Bhowmik, R. Sarkar, and M. Nasipuri, "Handwritten word recognition using lottery ticket hypothesis based pruned CNN model: a new benchmark on CMATERdb2.1.2," *Neural Computing and Applications*, vol. 32, no. 18, pp. 15209–15220, Sep. 2020, doi: 10.1007/s00521-020-04872-0.

[18] M. Awni, M. I. Khalil, and H. M. Abbas, "Deep-learning ensemble for offline Arabic handwritten words recognition," in *2019 14th International Conference on Computer Engineering and Systems (ICCES),* Cairo, Egypt, 2019, pp. 40–45, doi: 10.1109/ICCES48960.2019.9068184.

[19] M. Li *et al.*, "TrOCR: transformer-based optical character recognition with pre-trained models," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023, pp. 13094–13102, doi: 10.1609/aaai.v37i11.26538.

[20] A. Mostafa *et al.*, "OCFormer: a transformer-based model for Arabic handwritten text recognition," in *2021 International Mobile, Intelligent, and Ubiquitous Computing Conference, MIUCC 2021*, 2021, pp. 182–186, doi: 10.1109/MIUCC52538.2021.9447608.

[21] K. Barrere, Y. Soullard, A. Lemaitre, and B. Coüasnon, "A light transformer-based architecture for handwritten text recognition," *Document Analysis Systems: 15th IAPR International Workshop, DAS 2022*, Cham: Springer, 2022, pp. 275–290, doi: 10.1007/978-3-031-06555-2_19.

[22] A. Vaswani *et al.*, "Attention is all you need," *Advances in Neural Information Processing Systems*, pp. 5999–6009, 2017, doi: 10.1201/9781003561460-19.

[23] A. Dosovitskiy *et al.*, "An image is worth 16x16 words: transformers for image recognition at scale," *2021 International Conference on Learning Representations*, 2021, pp. 1.21.

[24] D. Svozil, V. Kvasnicka, and J. Pospichal, "Introduction to multi-layer feed-forward neural networks," *Chemometrics and Intelligent Laboratory Systems*, vol. 39, no. 1, pp. 43–62, 1997, doi: 10.1016/S0169-7439(97)00061-0.

[25] A. Al-Sabaawi, H. M. Ibrahim, Z. M. Arkah, M. Al-Amidie, and L. Alzubaidi, "Amended convolutional neural network with global average pooling for image classification," in *Intelligent Systems Design and Applications*, Cham: Springer International Publishing, 2021, pp. 171–180, doi: 10.1007/978-3-030-71187-0_16.

BIOGRAPHIES OF AUTHORS

Anju Arangil Thazhath is a research scholar in the Department of Computer Science at Sullamussalam Science College, Areekode, Malappuram, Kerala, India. She is also working as an assistant professor in data science and analytics at the University of Calicut, Kerala. Her research interests include artificial intelligence, machine learning, deep learning, image processing, and handwritten document image analysis, with a focus on Malayalam language word recognition. She has been actively engaged in research and academic initiatives in the field of intelligent systems and pattern recognition. She can be contacted at email: anjuat@sscollege.ac.in.

Dr. Binu Poothakuzhiyil Chacko is an associate professor in the Department of Computer Science at Prajyoti Niketan College, affiliated with the University of Calicut, Kerala, India. He holds a doctor of philosophy (Ph.D.) in Computer Science from Kannur University, Kerala. He has extensive teaching and research experience and currently serves as an approved research guide in computer science. His academic and research interests include artificial intelligence, machine learning, data analytics, and computational intelligence. He has guided several postgraduate research projects and published papers in reputed national and international journals. He can be contacted at email: binupchacko@gmail.com.

Prof. (Dr.) Mohamed Basheer Kizhakke Parambath is the principal of AMAL College of Advanced Studies, Nilambur, Kerala, India. He holds an M.Phil. and Ph.D. in Computer Science from Bharathidasan University, Tamil Nadu, India. With extensive academic and administrative experience, he has demonstrated excellence in teaching and research supervision. His research interests include machine learning, internet of things (IoT), artificial intelligence, speech technology, and biometric security systems. He has actively guided several research projects and contributed to the advancement of emerging technologies through his publications and academic initiatives. He can be contacted at email: mbasheerkp@gmail.com.