

An Efficient Handwritten Digit Classification Framework Based on Dynamic Attention Pyramid Head, Coordinate Attention, and Adaptive EfficientNetB7 with RP-DOA Optimization

Sarvesh Kumar Soni¹, Atul Srivastava^{*2}, Satyasundara Mahapatra³

^{1,2} Amity School of Engineering & Technology, AUUP, Lucknow, India

³Department of CSE, Pranveer Singh Institute of Technology, Kanpur, India

¹sarvesh8685@gmail.com, ²asrivastava5@lko.amity.edu, ³satyasundara123@gmail.com

Abstract: In recent times, handwritten numeral classification has attracted enormous interest in computer vision, because recognizing the digits is critical due to a wide range of individual handwriting styles. In multiple language countries like India, there is confined research on identifying handwritten characters and numerical by using deep learning, particularly when comparing with other additional vernacular scripts. The variations in writing shape, size, style, and similarities of different characters show accurate classification, adding even more challenges and complexity. Deep structured neural networks have appeared as an advanced technology for automatic object image classification and character patterns. Although deep neural networks have provided better performance on large datasets comprising millions of raw images, using deep network models on smaller datasets remains a challenging task. Hence, it is crucial to tackle difficulties that arise in the traditional digit classification model. In order to rectify a few issues presented in the previous research works, a novel deep structured learning-based digit classification framework is developed in this work. In the initial phase, essential images required for the evaluation are gathered from standard resources. Next, the collected input images are fed into the digit classification phase. Here, Dynamic Attention Pyramid Head with Coordinate Attention Parallel Fusion-based Adaptive EfficientnetB7 (DAPH-CAPF-AENetB7) is employed to carry out the classification procedure. The parameters of DAPH-CAPF-AENetB7 are tuned through the Renovated Position-based Dollmaker Optimization Algorithm (RP-DOA), which supports achieving more accurate digit classification results. Finally, several experiments are performed in the developed framework over classical techniques to check the overall efficiency of the developed framework in different classes.

Keywords: Handwritten digit recognition; deep learning; EfficientNetB7; attention mechanism; hyperparameter

1. Introduction

In recent times, in the field of computer vision, Handwritten Character Recognition (HCR) has received considerable attention. The major challenge in identifying handwritten digit stems from the high intra-class variation in individuals' writing styles. In a country like India, research on HCR using deep learning techniques remains limited. A frequent constraint in this field is the scarcity of accessible public datasets. Significantly, real-world applications focus on reliability rather than precision. The three basic tasks include recognition, interpretation, and identification, including handwritten documents. A digit classifier is a computerized program designed to identify printed or handwritten numerical images. It uses deep learning, where the system learns patterns from various examples instead of describing each feature of the digits. The most recent approaches use neural networks known as Convolutional Neural Networks (CNNs). These popular models assess the curves and shapes within images to compare digits. Newly



introduced digit images can be accurately classified by the trained model. This system outperforms traditional approaches when it comes to reading license plates, postal codes, and bank check numbers. It provides improved precision and accuracy.

Before deep learning came into existence, digit categorization was performed using classical machine-learning techniques with feature extraction. Specialists devised features such as structural traits, statistical metrics, and geometric properties to represent the digits. Commonly used ML classifiers include Support Vector Machines (SVM), K-Nearest Neighbours (KNN), Decision Trees, and Random Forests. To identify the closest match, template matching compared the input digits with stored templates. Such techniques achieved acceptable recognition accuracy for small datasets but demonstrated several problems when applied to various and large-scale handwriting samples. They were less adaptable, and their reliance on handcrafted features affected their performance and adaptability, resulting in a significant decrease in performance in the presence of noise, irregular strokes, and variations in style. These shortcomings showed the need for data-driven, adaptive, and scalable approaches that paved the way for the rise of deep-learning methods.

Deep learning-powered digit classifiers provide multiple advantages. They eliminate the requirement for handcrafted features, minimize human bias, and provide high robustness to illumination, handwriting variations, lighting changes, and noise. They can also effectively scale to large datasets. However, deep learning models require vast amounts of labelled data, which is resource-intensive to obtain, and require considerable computational resources for training and deployment. They may also be vulnerable to adversarial or noise attacks, and overfitting can occur without extensive training data. Furthermore, latency and energy requirements are limiting factors for embedded or mobile deployment. Research is underway to address these challenges through possible solutions. In this context, the dependency on labelled data can be reduced by using pre-trained models or unlabelled data through semi-supervised learning and transfer learning. Data augmentation methods, such as adding noise and applying geometric transformations, can address overfitting and increase data variety. Knowledge distillation and pruning/quantization can reduce the computational demands of models and enable their effective use in low-resource environments. Defensive distillation and adversarial training methods can enhance robustness against adversarial perturbations and noise.

Contribution: The proposed scheme is a hybrid scheme involving adaptive optimization, deep learning, and attention for an effective handwritten digit recognition system. This combination provides high precision, stability, and consistency in handwriting recognition. It can be used for archival document automation, education, and banking applications, as it can process documents in real time.

- To strengthen the discrimination and robustness of the features, DAPH-CAPF-AENetB7 is added to the framework. Dynamic Attention Pyramid Head (DAPH) adaptively changes the receptive fields to recognize digits with different shapes and orientations and extracts multiscale features. The Coordinate Attention (CA) mechanism captures positional information without discarding spatial information to direct the network's attention towards important digit locations and minimize distraction from noise. The revised EfficientNetB7 backbone is designed to leverage fine-grained and abstract features in parallel for accurate classification. The proposed architecture improves accuracy and offers faster inference and scalability, making it suitable for implementation in real-time handwritten digit recognition systems.
- Dynamic parameter tuning is introduced through RP-DOA to further optimize the model's training and adaptability using creative construction techniques. The method is used to adjust hyperparameters such as the learning rate, number of neurons, and batch size. The algorithm mimics intelligent crafting and modular design during training. This modification is beneficial for the convergence rate, performance, and stability of the model on various datasets. The structure of the Doll Maker Optimization Algorithm (DOA) provides a balance between exploration and exploitation. The network can perform efficiently to achieve the global optimum. The flexibility of the framework allows the development of real-time applications and enhances the power and scalability of the handwritten digit categorization framework in real-world applications.

2. Literature survey

In 2023, Goel and Ganatra [1] have resolved these challenges to segregate the handwritten Gujarati numerals from numerical ranges from 0 to 9 by utilizing three deep transfer-based learning instances. They suggested transfer learning methods using ten pre-trained Convolution Neural Network (CNN) architectures, to identify the highest performance of the model by fine-tuning and freezing weight parameters. These experiments were performed on a handwritten Gujarati digit dataset, which comprised 8000 raw images with data augmentation. Extensive experiments, including evaluation and testing, were carried out using different performance evaluation matrices to achieve the robustness and classification precision of the model. In 2022, Rasheed et al. [2] have implemented a classification architecture for automated handwritten recognition of Urdu digits and characters with enhanced recognition accuracy by using pre-trained and transfer learning using an optimized AlexNet for extracting classification and features, and a pre-trained AlexNet-CNN model integrated with a Support Vector Machine (SVM) classifier. They configured network hyperparameters to attain higher accuracy, and data augmentation was designed and implemented to avoid over-fitting. The developed research based on optimized network outperformed the other related state-of-the-art studies, thereby accomplishing a classification precision of 97.08%, 94.92%, and 98.21% for Urdu characters, hybrid datasets, and digits, respectively. In 2020, Weldegebriel et al. [3] have introduced the Handwritten Ethiopian Character Recognition (HECR) dataset to train the model, which contained 28×28 pixel images with more than one RGB colour pen image written with different colored pens. The dataset incorporates Ethiopian scripts (Fidel), combining tonal tools, numerals, special characters, and punctuation used in Eritrean and Ethiopian texts. A hybrid classification model integrating XGBoost was used for classification, and CNN for feature extraction. The result showcased that, while using the XGBoost as a classifier for the model, it excelled the conventional fully connected CNN layer regarding error rate. In 2021, Kusetogullari et al. [4] have proposed an unconventional deep structured learning approach named DIGINET and established a large-scale dataset of digits known as DIDA, which helps in detecting and recognizing handwritten digits in old documents. Utilizing the digitized images that were collected from Swedish handwritten records, the DIDA dataset was evolved, which was written by different priests with various handwriting styles. DIGINET comprises two significant components, named DIGINET-rec used in digit recognition and DIGINET-dect used in digit detection. DIGINET-rec operated the detected digits using three personalized CNN architectures by combining the outputs using a voting procedure for the final recognition. DIGINET-dect used three residual units, each individual unit with three CNN layers, and integrates a YOLO-based detection technique at two phases. In 2020, Sharmin et al. [5] have introduced a deep structured learning method for differentiating the Bengali spoken digits. It comprised parameters such as dialects, gender, and age groups into consideration, and the CNN-based proposed approach achieved an accuracy of more than 98%. The model resulted in better accuracy, improving its potential and robustness in speech recognition systems developed for diverse users for real-world applications. In 2021, Haghighi and Omranpour [6] have developed a new method of stacking ensemble classifier which relied on the Bidirectional Long-Short Term Memory (BLSTM) and the CNN. As an innovation in the model, the probability vector of the image was fed as input to the meta-classifier. The strength of the model was its learning ability from vectors and arrays, which makes it effective and efficient as a meta-classifier while using the output of the probability vector of a primary model and achieved a remarkable accuracy. In 2020, Ahlawat and Choudhary [7] suggested a hybrid model of advanced SVM and CNN for recognizing handwritten digits with MNIST dataset. In this work, the hybrid system which was constructed combining the essential aspects of two separate categorization methods was used. One component was a binary classifier and the other automatically extracted important features from the input data. The experimental results reveal the great performance of the system, being able to recognize the 99.280% of the handwritten digit dataset. In 2020, Zada and Ullah [8] have proposed a CNN-based isolated digits recognition system in the Pashto language. The database included Pashto digits ranging from 0 to 9, represented by 50 utterances for each digit. Twenty Mel-Frequency Cepstral Coefficients (MFCC) features were collected for individual digits and given as input to the CNN. The proposed network system included 4 convolutional layers, in which every layer underwent max pooling and ReLU activation layers. The network was pre-trained on the 50% data, and the remaining data was utilized for the testing phase. Finally, the system achieved a precision of 84.17% on testing and showed 7.32% improvement over the conventional approaches

3. Methods

3.1 Proposed DAPH-CAPF-AENetB7 Framework

The main benefit of the proposed structure is that it can distinguish the handwritten digits in different writing styles. A few advanced parts are embedded to enhance the learning capability of features and the adaptability of the model. Furthermore, the proposed solution is more scalable and can be easily used in large-scale real-world applications. In addition, it has the advantage of better accuracy and robustness in training and inference with the help of optimization techniques. It is important to develop a well-developed framework for categorization of digits using DAPH-CAPF-AENetB7 to enable a greater scalability and accuracy of recognition of handwritten digits. To achieve the best results, it is recommended to employ several deep learning networks. The CAPF takes into account the whole page content and the spatial context information for the important features, and the DAPF is able to highlight important features in handwritten numbers, thereby understanding the small differences. The EfficientNetB7 model enhances the capability of the system to be able to process complex handwriting variants efficiently. Here the RP-DOA modifies the critical parameters to enhance the overall categorization outcome. The RP-DOA algorithm optimizes the following parameters: hidden neuron count that improves the ability of the model to study intricate features and enhances accuracy, learning rate that controls the rate of adaptation of the model and ensures efficient convergence, and steps per epoch that controls the total number of iterations to achieve weight updates, and enhances the entire training process. Optimizing these parameters increases the framework's efficiency, robustness, and accuracy, which resulted in better results on automated handwritten digit classification tasks within a range of handwriting styles. The schematic representation of the overall architecture of the developed digit classification model is illustrated in Fig. 2(a).

3.2 Dataset

Collection of Handwritten Digit Images

This section describes the dataset utilized in the proposed system for precise digit classification for handwritten documents. The dataset can be freely accessed from Kaggle. This dataset contains a popular Modified National Institute of Standards and Technology (MNIST) dataset of 70.10k files, with a memory space of 39.78 MB. It comprises 70,950 images. The collected input image is represented in Fig.1 depicts the sample images of a digit.



Fig. 1 Sample Images of Handwritten Digits

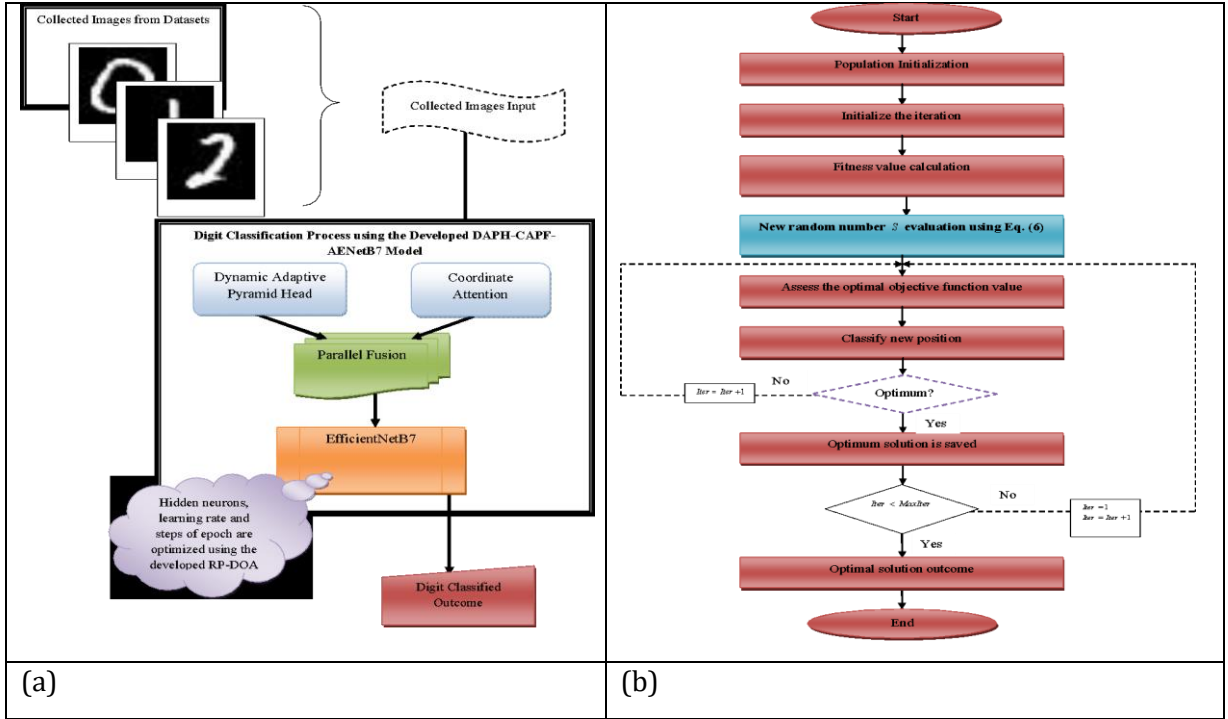


Fig. 2 (a) Overall Architecture of the Developed Digit Classification Model (b) Flow Chart Proposed RP-DOA Optimization Approach

3.3 RP-DOA Optimization

An innovative RP-DOA algorithm has been presented to offer an effective way to address parameter-related problems that come up during the classification stage. The objective of this strategy is to particularly address the shortcomings of the current DOA in terms of efficiently maximizing parameter configurations.

Novelty: The existing DOA [26] algorithm has been improved to provide a high-performance optimization method named RP-DOA. This improvement entails a calculated change to the original algorithm's random number-generating procedure. The proposed optimization algorithm can efficiently manage the exploration and exploitation phases, leading to more accurate and consistent optimization performance. As a result, there is less overfitting, increased detection accuracy, and more dependable parameter modifications. Thus, the proposed RP-DOA substantially enhances the detection technique's accuracy and robustness.

The RP-DOA is a human-inspired metaheuristic approach that initiates the adaptive behavior of stable systems by modifying their structural configurations, similar to how a dollmaker optimizes the model's components to achieve precision and balance. It is efficient in computing complex and different optimization problems by managing the search agents towards the optimal areas of the solution space. It describes consistent convergence behavior among high-dimensional, unimodal, and multimodal search spaces. Furthermore, RP-DOA is highly flexible and can easily be hybridized with other conventional algorithms, improving its applicability to specific problems in the domain. These strengths assure the proposed RP-DOA to outperform many existing optimization approaches in areas such as engineering design, parameter optimization, machine learning, and energy optimization.

Initially, to model the movement mathematically, a new position is calculated for each orangutan by using Eq. (1).

$$\chi_{c,d}^{q1} = \chi_{c,d} + n \cdot (Q_{c,d} \cdot S \cdot \chi_{c,d}) \quad (1)$$

Here, the term $\chi_{c,d}^{q1}$ is the newly chosen doll-making position of the c^{th} orangutan related to the initial phase of the RP-DOA algorithm, $\chi_{c,d}^{q1}$ which is its d^{th} dimension, n a random number drawn from the interval $[0,1]$ and $Q_{c,d}$ implies the d^{th} dimension of the input digit for the c^{th} iteration orangutan. This movement allows the orangutan to study different regions within the search area, which results in an advancement in a high-quality optimization target, potentially, and it is illustrated in Eq. (2).

$$\chi_{c,d}^{q1} = \chi_{c,d} + \frac{CurFt}{CurFt+WtFt} (Q_{c,d} - S \cdot \chi_{c,d}) \quad (2)$$

Here, the terms $CurFt$ represent the current fit, and $WtFt$ represent the worst fit. The RP-DOA algorithm is proposed to efficiently find optimal solutions by effectively maintaining a dynamic balance between exploration and exploitation stages. It enhances the detection system's performance by accurately aligning with both system and user needs, which leads to improved clarity, automation, and consistency in the initial stages of detection. The updating of the position according to the random number provides an automatic and systematic way of assisting classification tasks, making it adaptable and efficient for applications such as handwritten digit document classification. The algorithm guides optimal solutions progressively toward optimal regions of the search space, along with maintaining population diversity to avoid local minima. Its lightweight structure allows for increased convergence and requires minimal parameter optimization, which also permits adaptability towards unimodal and multimodal objective functions. The schematic flowchart representation of the proposed RP-DOA optimization approach is illustrated in Fig. 2(b).

a. 3.4. Basic EfficientnetB7

EfficientNetB7 [31] is renowned for its precision and performance efficiency, especially in classification applications. Each of its seven successive blocks was constructed using Mobile Reversed Bottleneck Convolution (MRBC) layers with different strides, channel depths, and filter sizes. The compound scaling method, which uses a compound coefficient φ to uniformly scale the network's depth denoted as T , width denoted as B , and, resolution denoted as U . These play a significant role in EfficientNetB7. This approach guarantees a balanced model expansion to improve efficiency and accuracy. The mathematical representation of the scaling approach is given in Eqs. (3), (4), and (5).

$$T = \beta^\varphi \quad (3)$$

$$B = \mu^\varphi \quad (4)$$

$$U = \chi^\varphi \quad (5)$$

A configurable scaling factor is $\varphi \geq 1$. EfficientNetB7 is a favoured option for transfer computation and medical imaging classification tasks because of its formulation, which preserves a well-proportioned model that can flexibly adapt to changing hardware restrictions while maintaining high accuracy.

$$\beta = \left(\frac{1}{s}\right) \times \sum Y \quad (6)$$

$$\eta^2 = \left(\frac{1}{s}\right) \times \sum (Y - \kappa)^2 \quad (7)$$

Here, in Eq. (6) and (7), the mean and variance for the first mini-batch are computed in which the term β denotes the mean, the term η^2 signifies the variance, and the term s indicates the count of instances in the mini-batch. To standardise the activation to have a mean of zero and a variance of one, the following Eq. (8) is formulated.

$$Y_{hat} = (Y - \beta) / \sqrt{(\eta^2 + \varepsilon)} \quad (8)$$

Here, the term ϵ represents a small constant that ensures numerical stability. In addition, scaling and shifting techniques are implemented to grant the neural network the ability to adjust normalized activations for particular tasks, achieved by incorporating learnable parameters beta and gamma.

$$H = \chi \times Y_{hat} + \gamma \quad (9)$$

In Eq. (9), γ denote the learnable parameters that get modified throughout the training process via backpropagation, enabling the network to fine-tune the scaling and shifting factors for every layer.

4. Adaptive Deep Learning Model for Digit Classification in Handwritten Documents and its Performance using Enhanced Optimization Algorithm

4.1 DAPH-CAPF-AENetB7-based Digit Classification Model in Handwritten Documents

The DAPH-CAPF-AENetB7-based digit classification model, shown in Fig. 3, is an advanced framework proposed to efficiently recognize and classify digits in handwritten documents. This model incorporates several techniques, which include DAPH, CAPF, and AENetB7, to optimize the classification phase.

Novelty: The digit classification model based on DAPH-CAPF-AENetB7 starts with the data collection, which involves the collection of handwritten documents from various data sources. The acquired image basically travels through the initial MBConv (Mobile Inverted Bottleneck Convolution) layers of EfficientNet-B7 which steadily reduces spatial dimensions while enhances the feature representation quality. At a carefully selected MBConv layer, the generated feature map is split and fed into two attention modules such as the DAPH and the CA [32] module. By combining pyramid pooling, feature pyramids, and dynamic attention techniques, the DAPH [30] module successfully captures both spatial and contextual semantics, enhancing the network's capacity to localize discriminating data through spatial encoding. This enables the model to highlight significant high-level and fine-grained areas in the digitised images, which is essential for precise categorization. DAPH improves the classification of handwritten digits. The given feature pyramid obtained by coordinate-wise summation operations, as $Q \in S^{L \times H \times M \times O}$, where O denotes the obtained feature pyramid level, Dynamic Attention Module (DAM) [34] performs scale-conscious attention to combine the features dynamically from various levels to extract the semantic importance as represented in Eq. (10).

$$g(Q) = (0, (1, (g(\frac{1}{LH \times M} \sum_{LH, M} Q) + 1)/2)).Q \quad (10)$$

DAM learns to focus on differentiating areas that already exist in spatial areas and are disturbed by feature levels by enforcing spatial-aware attention further. This consists of the two following steps. Firstly, the DAM utilizes the deformable convolution to generate the sparse attention in learning. Then, the similar spatial features among various clustering levels at similar spatial locations are divided, which can be calculated as given in Eq. (11).

$$g(Q) = \frac{1}{M} \sum_{m=1}^M \sum_{l=1}^L v_{m,l} \cdot Q(m; q_l + \Delta q_l; m) \cdot \Delta n_l \cdot Q \quad (11)$$

Here, the term denotes the sparse sampling area number, $q_l + \Delta q_l$ denotes a changed location by the auto-learned offset Δq_l to concentrate on a differentiative region, Δn_l represents a learned significance scalar at the location Δn_l , and Δq_l is studied by the median level of Q the input feature.

By breaking down channel attention into unidirectional components, the CA module simultaneously encodes exact positional dependencies, boosting spatial encoding and the network's capacity to localize discriminative information. A united attention-enhanced feature representation is subsequently produced by fusing the outcomes of these two attention modules through Parallel Fusion and fed into the MBConv layer. The EfficientNetB7 backbone then passes this combined feature map into further MBConv layers for additional processing and improvement, allowing the model to produce reliable and precise predictions on tasks involving the classification of handwritten digits. Furthermore, the RP-DOA is used to tune the network parameters for hidden neuron, learning rate and steps per epoch to increase the generalization and robustness of the final model in the decision-making process. The

effective exploration and exploitation of the solution space is achieved with position-based encoding and adaptive mutation mechanisms in RP-DOA. Then the resultant classification output is produced ensuring robustness to variance, high accuracy and excellent generalization, especially for handwritten digit document images typically found in the real world.

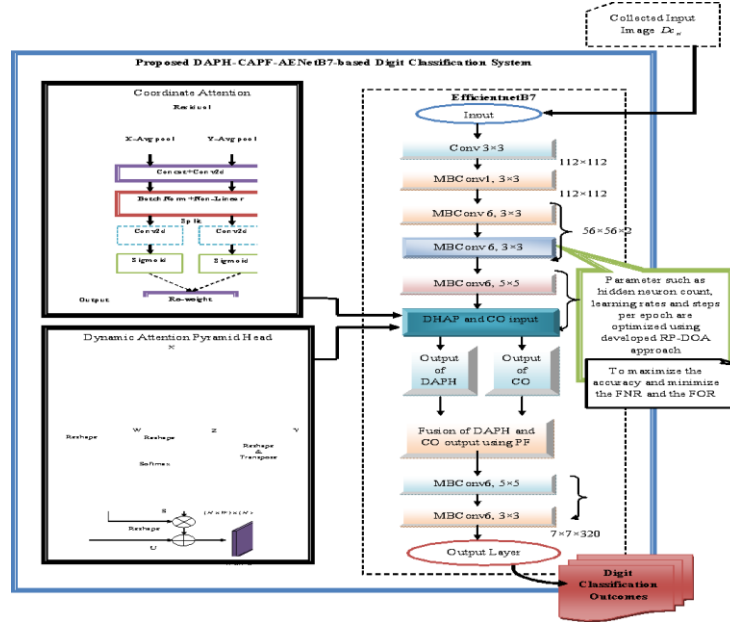


Fig. 3 Schematic Visualisation of the Proposed DAPH-CAPF-AENetB7-based Digit Classification Model in Handwritten Documents

4.2 Objective Function

For the digit classification of handwritten documents, the objective function of the proposed model is given in Eq. (10). The use of the objective function plays a very important role in enhancing the effectiveness of the system as it allows for easy categorization of digits in the design of the proposed system. It calculates the classification objective using the digit by maximizing accuracy, and minimizing False Negative Rate (FNR) and False Omission Rate (FOR), respectively. The parameters named hidden neuron count, learning rate, and number of epochs are optimized using RP-DOA. While minimizing FNR, it reduces missed digit predictions and minimizing FOR, avoids the incorrectly classified non-digits as digits, maximizing accuracy in digit classification guarantees precise recognition of both miswritten and legitimate digits. Overall classification performance is optimized by this balance.

$$OFn_f^G = \arg \min_{\{Nl_n^{neu}, LnRt_n^{rn}, EpOh_h^o\}} \left(\frac{1}{NQ} + Qn + D_{occ} \right) \quad (12)$$

Here, in Eq. (10), the term OFn_f^G denotes the objective function, the terms Nl_n^{neu} , $LnRt_n^{rn}$ and $EpOh_h^o$ represent the count of hidden neurons, learning rate, and number of epochs in RAN that range between $[5 - 255]$, $[0.01 - 0.99]$, and $[100 - 500]$ respectively. The term NQ represents the accuracy, the term Qn denotes FNR, and the term D_{occ} represents FOR.

$$NQ = \frac{c+Q}{c+Q+M+S} \quad (13)$$

$$Qn = \frac{S}{c+S} \quad (14)$$

$$D_{occ} = \frac{S}{s+Q} \quad (15)$$

Here, the terms C, Q, M, S represent the true positive, true negative, false positive, and false negative, respectively, in Eqs. (13), (14), and (15).

b. Performance of the DAPH-CAPF-AENetB7 with RP-DOA in Digit Classification

Implementing data entry along with information retrieval procedures in a variety of industries, including banking, administration, and education, depends heavily on accurate digit categorization in handwritten documents. However, conventional methods of classification face significant challenges due to the variety of distortions, styles, and noises seen in handwritten inputs. In order to get over these restrictions, the suggested framework combines a cutting-edge optimization technique called RP-DOA with a potent deep learning model called DAPH-CAPF-AENetB7 to greatly improve classification performance. The goal of the model DAPH-CAPF-AENetB7 is to gain rich context-aware properties from complex handwritten digits. The DAPH module always looks at important visual cues at varying scales, this may allow the model to focus on important stroke patterns and curvature in handwritten digits. The spatial and contextual information of the CAPF mechanism can improve the discriminating power of the feature maps and boost the spatial attention. The backbone network AENetB7 dynamically adjusts the depth, width, and clarity to achieve the optimal trade-off between accuracy and computation efficiency. The RP-DOA method is used to optimize the model for better performance. This metaheuristic optimization technique smartly tunes essential hyperparameters such as number of hidden neurons, learning rate and steps per epoch. RP-DOA lowers the training loss, avoids over-fitting and ensures optimal convergence by effectively searching the optimal space. This hybrid approach is suitable for real-world handwritten document digitization since it minimizes false predictions, maximizes the total classification precision, and provides good generalization for different handwriting styles.

5. Results and Discussion

a. Simulation Set-up

Python was used to develop digit classification for handwritten documents mechanisms. The optimization annotations are slightly changed to incorporate a population of size 10, chromosomal length of 3, as well as a maximum iteration of 50, to effectively classify digits in handwritten documents. The performance of proposed model was compared with different existing network models, namely AlexNet-SVM [2], Convolutional Neural Networks XGBoost (CNN-XGBoost) [3], CNN-BLSTM [6], and EfficientNetB7 [31], and optimization algorithms, such as Crystal Structure Algorithm (CSA) [27], Fennec Fox Optimization (FFO) [28], Lyrebird Optimization Algorithm (LOA) [29], and Dollmaker Optimization Algorithm (DOA) [26] were utilized for the productive comparison with the proposed digit classification for handwritten documents.

b. Performance Metrics

This section demonstrates the performance metrics used during validation.

$$(a) \text{ Sensitivity: } Tr = \frac{C}{C+S} \quad (16)$$

$$(b) \text{ Specificity: } Cm = \frac{Q}{Q+M} \quad (17)$$

$$(c) \text{ Precision: } Ir = \frac{C}{C+M} \quad (18)$$

$$(d) \text{ False Positive Rate (FPR): } Ut = \frac{M}{M+Q} \quad (19)$$

$$(e) \text{ Negative Predictive Value (NPV): } Sd_F = \frac{Q}{Q+S} \quad (20)$$

$$(f) \text{ False Discovery Rate (FDR): } Rv_i = \frac{M}{C+M} \quad (21)$$

(g) F1 Score:

(22)

$$Rs_{sc} = \frac{2 \times C}{(2 \times C) + M + S}$$

(h) Matthews Correlation Coefficient (MCC): $Tb_w =$

$$\frac{(C \times Q) - (M \times S)}{\sqrt{(C+M) \times (C+S) \times (Q+M) \times (Q+S)}} \quad (23)$$

(i) Prevalence Threshold (PT):

$$Su_t = \frac{\sqrt{(Ut \times Ut)} - Ut}{Tr - Ut} \quad (24)$$

(j) Critical Success Index (CSI):

$$CSi = \frac{C}{C + Q + M} \quad (25)$$

(k) Balanced Accuracy (BA):

$$Rp = \frac{Tr + Cm}{2} \quad (26)$$

(l) Fowlkes–Mallows Index (FM):

$$Mg = \sqrt{Ir \times Tr} \quad (27)$$

(m) Bookmaker Informedness (BM):

$$Bh = Tr + Cm - 1 \quad (28)$$

(n) Markedness (MK):

$$Sk = Ir + Sd_f - 1 \quad (29)$$

(o) Positive Likelihood Ratio (LR+):

$$LnT = \frac{Tr}{Ut} \quad (30)$$

(p) Negative Likelihood Ratio (LR-):

$$LmT = \frac{Qn}{Cm} \quad (31)$$

(q) Diagnostic Odds Ratio (DOR):

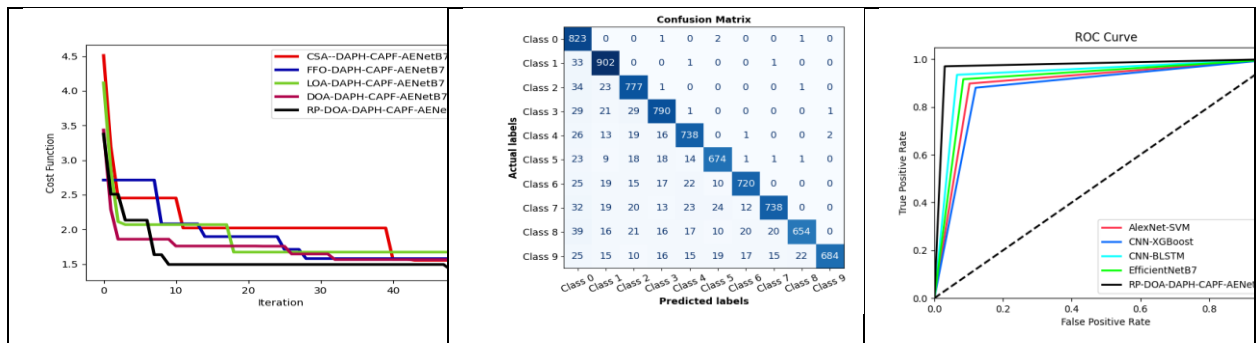
$$OT = \frac{LnT}{LmT} \quad (32)$$

(r) Prevalence:

$$Un_y = \frac{Q+S}{C+Q+K+S} \quad (33)$$

C. Cost Function Evaluation of the Developed Model

The cost function analysis of the developed digit classification model is illustrated in Fig. 4 (a). The model operates better with digits and characterizes its resilience, according to the computation of the various cost functions. Iterative training greatly increases the model's stability and allows the digit classification model to handle and examine the intricate and complex properties of digits. CSA-DAPH-CAPF-AENetB7, FFO-DAPH-CAPF-AENetB7, LOA-DAPH-CAPF-AENetB7, and DOA-DAPH-CAPF-AENetB7 were compared to the suggested model for a dataset. The comparative results represent the high performance of the proposed model with the values of 23.333% of CSA-DAPH-CAPF-AENetB7, 25.454% of FFO-DAPH-CAPF-AENetB7, 15.853% of LOA-DAPH-CAPF-AENetB7, and 1.470% of DOA-DAPH-CAPF-AENetB7 in digit classification of handwritten documents for a dataset. With values of 23.333% for CSA-DAPH-CAPF-AENetB7, 25.454% for FFO-DAPH-CAPF-AENetB7, 15.853% for LOA-DAPH-CAPF-AENetB7, and 1.470% for DOA-DAPH-CAPF-AENetB7 in the digit classification of handwritten documents for a dataset, the comparison findings demonstrate the outstanding performance of the proposed model. This supports and enhances the model in understanding complex handwritten patterns better across different writing styles.



(a)	(b)	(c)
-----	-----	-----

Fig. 4 Evaluation of (a) Cost Function (b) Confusion Matrix (c) ROC curve of Developed Model

D. Confusion Matrix Evaluation of the Developed Model

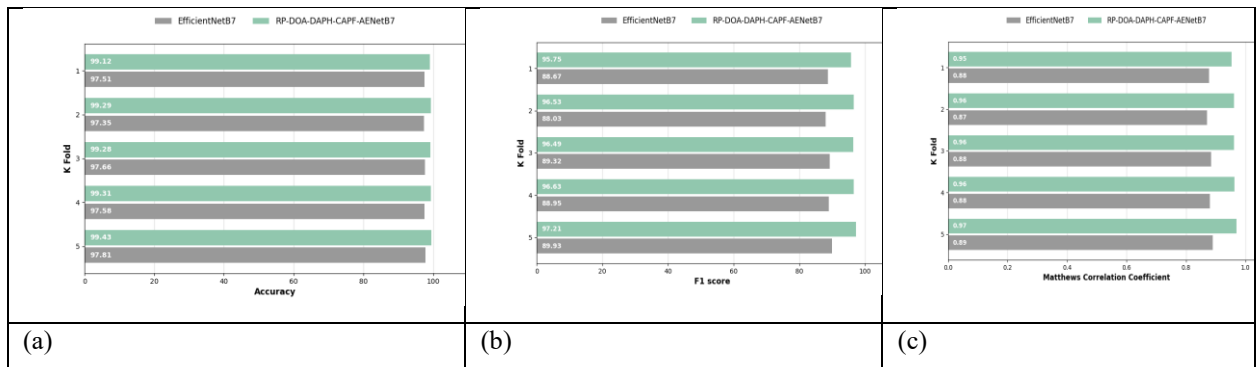
Fig. 4 (b) depicts the confusion matrix evaluation of the proposed model of a dataset in digit classification. A confusion matrix is a performance measurement tool that is used in classification tasks to compute how good a model predicts the class labels. The elements of the confusion matrix are presented in tabular form for each class to give an overview of the results. This helps you to understand overall accuracy as well as the accuracy of each class. The confusion matrix is essential for making sure that the proposed architecture is effective in developing an efficient digit classification system for handwritten documents with the help of DAPH-CAPF-AENetB7. It points out the errors in classification of handwritten digits with similar appearance and helps the researchers to improve their attention mechanisms and attention fusion strategies. The analysis of the matrix helps to identify the specific aspects that need to be addressed for more consistent and accurate digit recognition in cases of handwritten text.

E. ROC Evaluation of the developed model

Fig. 4 (c) shows the Receiver Operating Characteristic (ROC) evaluation of the proposed model on a dataset. The ROC analysis is provided to compute the degree of sensitivity and specificity, which enables the model's classification performance. The ROC validation demonstrates that the proposed model achieves a higher True Positive Rate (FPR) and FPR rate in digit classification for handwritten documents. On comparing with other conventional algorithms such as AlexNet-SVM, CNN-XGBoost, CNN-BLSTM, and EfficientNetB7, the proposed model outperformed by 15.429%, 18.072%, 12.643%, and 15.294% respectively. Thus, the proposed model achieves better than other models, outperforming with strong TPR and minimal FPR values, demonstrating robustness and effectiveness of the model with higher accuracy.

F. K-Fold-based Evaluation of the Developed Model

Fig. 5 illustrates the evaluation of the developed model for a dataset using a K-fold analysis. The suggested model's generalization and dependability are effectively facilitated by K-fold analysis. Dividing the datasets into several folds for training and validation minimizes bias and avoids overfitting problems. It also displays a comprehensive estimation of performance across multiple datasets. For different dataset folds, the developed algorithm displayed 1.610% higher MCC than the conventional approach when comparing the outcomes of the suggested DAPH-CAPF-AENetB7 with EfficientNetB7. Hence, the developed model increases the computational efficiency while maintaining the accuracy as high, which ensures rapid processing using a few parameters, which makes the model better for source-limited and real-time applications.



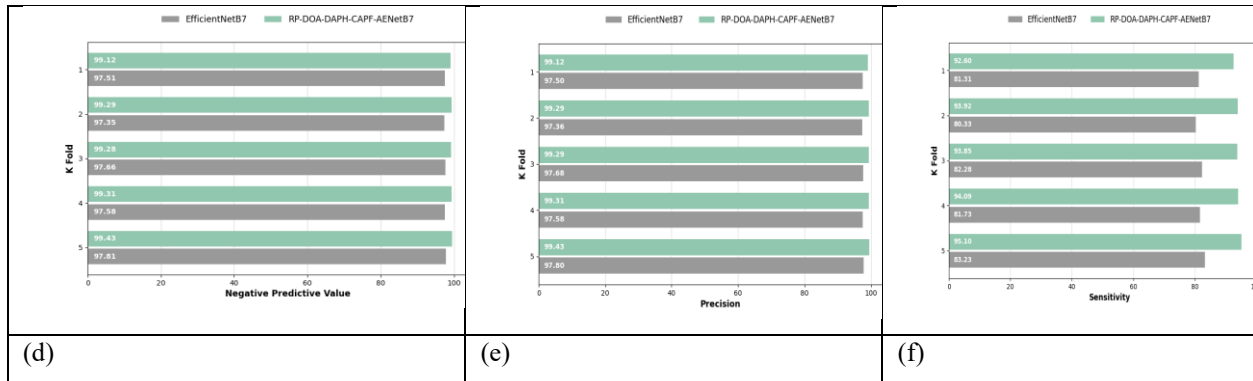
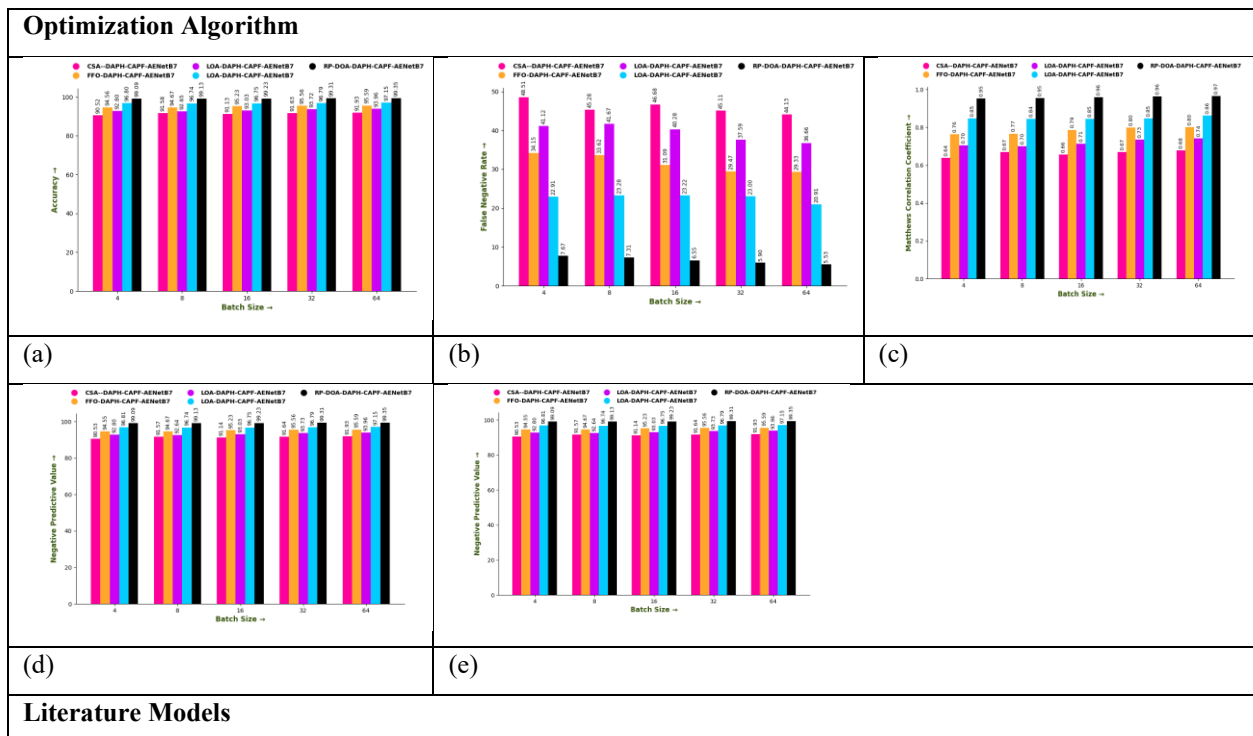


Fig.5 K-Fold Evaluation of the Developed Model based on (a) Accuracy, (b) F1 Score, (c) MCC, (d) NPV, (e) Precision and (d) Sensitivity

G. Batch Size-based Classification Analysis of the Proposed Model

Fig. 6 shows the suggested model's batch size-based classification over the optimization algorithm and literature models. It is required to calculate the produced digit classification model's reliability. It explains the model's robustness, accuracy, and generalization capabilities, allowing it to outperform other traditional methods. For CSA-DAPH-CAPF-AENetB7, FFO-DAPH-CAPF-AENetB7, LOA-DAPH-CAPF-AENetB7, and DOA-DAPH-CAPF-AENetB7, the NPV of current methods against the developed model for a dataset is 8.071%, 3.933%, and 5.736%, respectively. Based on this assessment, the model showcased high accuracy, FNR, MCC, NPV, and Precision. The research demonstrates that the suggested model exhibits strong class division, high accuracy, and fewer misclassifications.



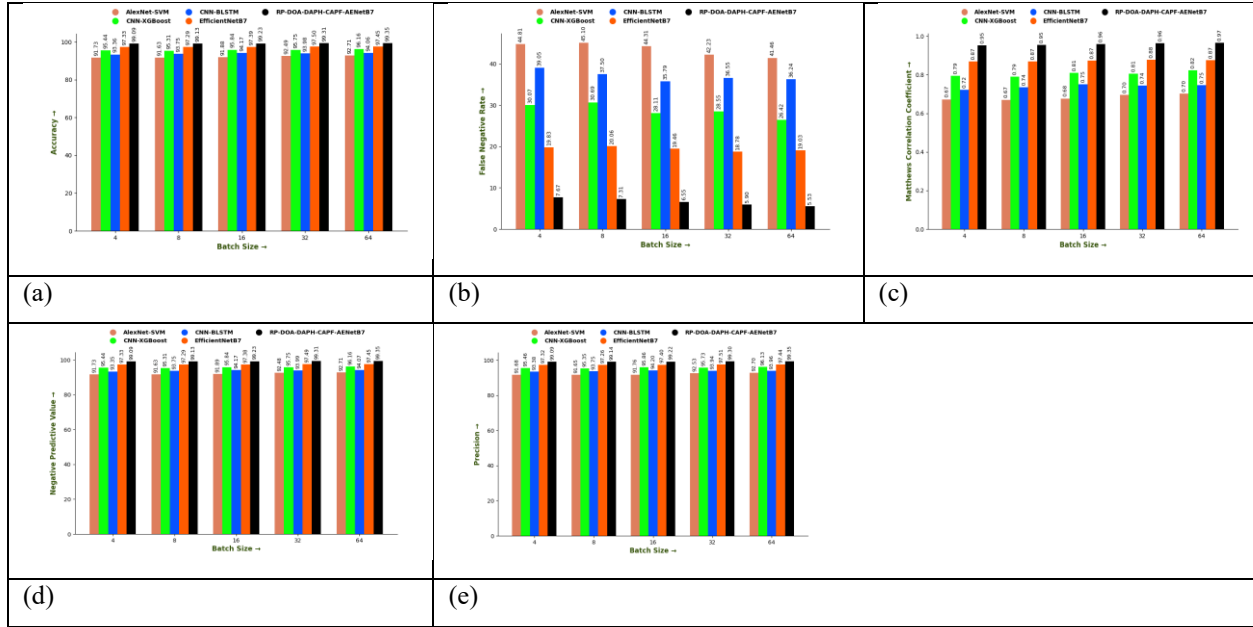


Fig. 6 Classification Analysis of the Proposed Model over Different Optimization and Literature Models

6. Conclusion

This work presented a robust handwritten digit classification model using the DAPH-CAPF-AENetB7 architecture, designed to improve accuracy in complex and low-resource numeral datasets. It is ideally suited for useful, real-world digit recognition systems because of its design, which guaranteed both feature diversity and training efficiency. This paper developed a novel handwritten digit classification model using the DAPH-CAPF-AENetB7 architecture, which incorporated. The process starts with the collection of handwritten numeral images from benchmark datasets. These input images are passed through the DAPH-CAPF-AENetB7 network, where DAPH modules and CA mechanisms focus on important spatial and channel-wise features. The parallel fusion helped in integrating varied feature representations and the RP-DOA method optimizes the crucial hyperparameters, namely learning rate, neuron count, and steps of epochs, ensuring effective convergence and avoiding local minima, thus delivering optimized classification performance. The proposed model achieved 5.348%, 4.770%, 8.825% and 6.449% higher precision rate compared to the existing optimization algorithms, such as CSA-DAPH-CAPF-AENetB7, FFO-DAPH-CAPF-AENetB7, LOA-DAPH-CAPF-AENetB7 and DOA-DAPH-CAPF-AENetB7. This enhances the robustness and accuracy, particularly on diverse small datasets. This made it very suitable for real-time applications where the variances in writing are large and the labeled data is extremely restricted.

Limitations and Future Scopes: The suggested model achieves improved efficiency. However, it is influenced by the limited availability of large-scale multilingual numeric datasets and the variation of handwriting styles. Its dependence on EfficientNetB7 increases the computational complexity and makes it less suited for low power systems. Future work can be directed on producing lightweight models for mobile deployment and more intense regional datasets. Future work can explore semi/self-supervised learning, ensemble approaches and transformer-based modules to lessen reliance on big labeled datasets and improve accuracy.

CRedit Author Contribution Statement:

Sarvesh Kumar Soni: Investigation, Data Curation, Formal Analysis, Visualization, Writing – Original Draft.
Atul Srivastava: Conceptualization, Methodology, Software, Validation, Formal Analysis, Supervision, Project Administration, Writing – Review & Editing.
Satyasundara Mahapatra: Supervision, Validation, and Writing – Review & Editing

Conflicts of Interest: The authors declare that they have no conflict of interest.

Data Availability Statement: The dataset used in this study is publicly available through Kaggle.

Use of AI Tools: Generative artificial intelligence tools were used to assist with language editing and manuscript presentation. The authors reviewed and verified the content and remain responsible for the accuracy and final content of the manuscript.

Ethics Approval: Ethical approval was not required for this study because it used a publicly available dataset and did not involve direct recruitment of human participants.

Acknowledgements: The authors have no acknowledgements to declare.

Funding: This research received no external funding.

REFERENCES

1. P. Goel and A. Ganatra, "Handwritten Gujarati Numerals Classification Based on Deep Convolution Neural Networks Using Transfer Learning Scenarios," *IEEE Access*, vol. 11, pp. 20202–20215, 2023.
2. A. Rasheed, N. Ali, B. Zafar, A. Shabbir, M. Sajid, and M. T. Mahmood, "Handwritten Urdu Characters and Digits Recognition Using Transfer Learning and Augmentation With AlexNet," *IEEE Access*, vol. 10, pp. 102629–102645, 2022.
3. H. T. Weldegebriel, H. Liu, A. U. Haq, E. Bugingo, and D. Zhang, "A New Hybrid Convolutional Neural Network and eXtreme Gradient Boosting Classifier for Recognizing Handwritten Ethiopian Characters," *IEEE Access*, vol. 8, pp. 17804–17818, 2020.
4. H. Kusotogullari, A. Yavariabdi, J. Hall, and N. Lavesson, "DIGITNET: A Deep Handwritten Digit Detection and Recognition Method Using New Historical Handwritten Digit Dataset," *Big Data Research*, vol. 23, no. 100182, 2021.
5. R. Sharmin, S. K. Rahut, and M. R. Huq, "Bengali Spoken Digit Classification: A Deep Learning Approach Using Convolutional Neural Network," *Procedia Computer Science*, vol. 171, pp. 1381–1388, 2020.
6. F. Haghighi and H. Omranpour, "Stacking ensemble model of deep learning and its application to Persian/Arabic handwritten digits recognition," *Knowledge-Based Systems*, vol. 220, no. 106940, 2021.
7. S. Ahlawat and A. Choudhary, "Hybrid CNN-SVM Classifier for Handwritten Digit Recognition," *Procedia Computer Science*, vol. 167, pp. 2554–2560, 2020.
8. B. Zada and R. Ullah, "Pashto isolated digits recognition using deep convolutional neural network," *Heliyon*, vol. 6, no. 2, 2020.
9. H. Shi and D. Zhao, "License Plate Recognition System Based on Improved YOLOv5 and GRU," *IEEE Access*, vol. 11, pp. 10429–10439, 2023.
10. G. Zhao, W. Wang, X. Wang, X. Bao, H. Li, and M. Liu, "Incremental Recognition of Multi-Style Tibetan Character Based on Transfer Learning," *IEEE Access*, vol. 12, pp. 44190–44206, 2024.
11. A. Tourani, A. Shahbahrani, S. Soroori, S. Khazaei, and C. Y. Suen, "A Robust Deep Learning Approach for Automatic Iranian Vehicle License Plate Detection and Recognition for Surveillance Systems," *IEEE Access*, vol. 8, pp. 201317–201330, 2020.
12. M. R. Tonmoy, M. A. Adnan, A. K. Saha, M. F. Mridha, and N. Dey, "Descriptor: Multilingual Visual Font Recognition Dataset," *IEEE Data Descriptions*, vol. 1, pp. 8–12, 2024.
13. A. Fateh, R. T. Birgani, M. Fateh, and V. Abolghasemi, "Advancing Multilingual Handwritten Numeral Recognition With Attention-Driven Transfer Learning," *IEEE Access*, vol. 12, pp. 41381–41395, 2024.
14. K. U. U. Rehman and Y. D. Khan, "A Scale and Rotation Invariant Urdu Nastalique Ligature Recognition Using Cascade Forward Backpropagation Neural Network," *IEEE Access*, vol. 7, pp. 120648–120669, 2019.
15. S. M. Darwish and K. O. Elzoghaly, "An Enhanced Offline Printed Arabic OCR Model Based on Bio-Inspired Fuzzy Classifier," *IEEE Access*, vol. 8, pp. 117770–117781, 2020.
16. Y. Deng, G. Wang, C. Li, W. Wang, C. Zhang, and J. Tang, "Collaborative License Plate Recognition via Association Enhancement Network With Auxiliary Learning and a Unified Benchmark," *IEEE Transactions on Multimedia*, vol. 26, pp. 11402–11414, 2024.
17. T. Ebadi, I. Kukenys, W. N. Browne, and M. Zhang, "Human-Interpretable Feature Pattern Classification System Using Learning Classifier Systems," *Evolutionary Computation*, vol. 22, no. 4, pp. 629–650, Dec. 2014.
18. A. D. Le, T. Clanuwat, and A. Kitamoto, "A Human-Inspired Recognition System for Pre-Modern Japanese Historical Documents," *IEEE Access*, vol. 7, pp. 84163–84169, 2019.
19. G. Siebra Lopes, D. Clifte da Silva, A. W. Oliveira Rodrigues, and P. P. Reboucas Filho, "Recognition of Handwritten Digits Using the Signature Features and Optimum-Path Forest Classifier," *IEEE Latin America Transactions*, vol. 14, no. 5, pp. 2455–2460, May 2016.
20. H. Shi and D. Zhao, "License Plate Recognition System Based on Improved YOLOv5 and GRU," *IEEE Access*, vol. 11, pp. 10429–10439, 2023.
21. A. B. M. Ashikur Rahman, M. B. Hasan, S. Ahmed, and T. Ahmed, "Two Decades of Bengali Handwritten Digit Recognition: A Survey," *IEEE Access*, vol. 10, pp. 92597–92632, 2022.

22. T.-C. Hsu, Y.-H. Tsai, and W. Cheng-Chung Chu, "Enhancing Digit Recognition for Luminous Images in Edge Computing Through Transfer Learning With Robustness and Fault Tolerance," *IEEE Transactions on Reliability*, vol. 74, no. 2, pp. 2529–2537, June 2025.
23. A. F. De Sousa Neto, B. L. D. Bezerra, E. B. Lima, and A. H. Toselli, "HDSR-Flor: A Robust End-to-End System to Solve the Handwritten Digit String Recognition Problem in Real Complex Scenarios," *IEEE Access*, vol. 8, pp. 208543–208553, 2020.
24. A. Tigrini, F. Verdini, M. Scattolini, F. Barbarossa, and L. Burattini, "Handwritten Digits Recognition From sEMG: Electrodes Location and Feature Selection," *IEEE Access*, vol. 11, pp. 58006–58015, 2023.
25. R. Malhotra and M. T. Addis, "Handwritten Amharic Word Recognition With Additive Attention Mechanism," *IEEE Access*, vol. 12, pp. 114645–114657, 2024.
26. S. Al Omari, K. Kaabneh, I. Abufalahah, and K. Eguchi, "Dollmaker Optimization Algorithm: A Novel Human-Inspired Optimizer for Solving Optimization Problems," *International Journal of Intelligent Engineering and Systems*, vol. 17, no. 3, pp. 816–828, 2024.
27. A. Kadan, K. Ryczko, A. Wildman, R. Wang, A. Roitberg, and T. Yamazaki, "Accelerated Organic Crystal Structure Prediction With Genetic Algorithms and Machine Learning," *Journal of Chemical Theory and Computation*, vol. 19, no. 24, pp. 9388–9402, 2023.
28. Y. Shen, D. Li, and W. Wang, "Multi-Energy Load Prediction Method for Integrated Energy System Based on Fennec Fox Optimization Algorithm and Hybrid Kernel Extreme Learning Machine," *Entropy*, vol. 26, no. 8, p. 699, 2024.
29. M. Dehghani, G. Bektemyssova, Z. Montazeri, G. Shaikemelev, O. P. Malik, and G. Dhiman, "Lyrebird Optimization Algorithm: A New Bio-Inspired Metaheuristic Algorithm for Solving Optimization Problems," *Biomimetics*, vol. 8, no. 6, p. 507, 2023.
30. Y. Xu, S. Zhou, and Y. Huang, "Transformer-Based Model With Dynamic Attention Pyramid Head for Semantic Segmentation of VHR Remote Sensing Imagery," *Entropy*, vol. 24, no. 11, p. 1619, 2023.
31. H. M. T. Khushi, T. Masood, A. Jaffar, M. Rashid, and S. Akram, "Improved Multiclass Brain Tumor Detection via Customized Pretrained EfficientNetB7 Model," *IEEE Access*, vol. 11, pp. 117210–117230, 2023.
32. C. Kang, L. Jiao, K. Liu, Z. Liu, and R. Wang, "Precise Crop Pest Detection Based on Co-Ordinate-Attention-Based Feature Pyramid Module," *Insects*, vol. 16, no. 1, p. 103, 2025.
33. P. Ding, H. Qian, Y. Zhou, S. Yan, S. Feng, and S. Yu, "Real-Time Efficient Semantic Segmentation Network Based on Improved ASPP and Parallel Fusion Module in Complex Scenes," *Journal of Real-Time Image Processing*, vol. 20, no. 3, p. 41, 2023.
34. X. Li, J. Dong, J. Tang, and J. Pan, "DLGSANet: Lightweight Dynamic Local and Global Self-Attention Networks for Image Super-Resolution," *IEEE*, pp. 12792–12801, 2023.