Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 4, August 2026, pp. 3109∼3119
ISSN: 2302-9285, DOI: 10.11591/eei.v15i4.11113

Enhancing image classification accuracy with AL-CNN: a hybrid of AlexNet and LeNet architectures

Monika Chawla[1], Rashmi Agrawal[1], Bharat Bhushan[2]
[1]Department of Computer Application, School of Computer Application, Manav Rachna International Institute of Research and Studies, Faridabad, India
[2]School of Healthcare and Allied Sciences, GD Goenka University, Gurugram, India

Article Info

Article history:

Received Jul 27, 2025
Revised Apr 25, 2026
Accepted Jul 9, 2026

Keywords:

Adaptive learning convolutional neural network
CIFAR-10
Convolutional neural networks
Fashion-MNIST
Graph neural networks
Image classification
Recurrent neural networks

ABSTRACT

Image classification is a key application of computer vision with direct relevance to medical diagnostics, autonomous vehicles, and remote sensing. This paper discusses the use of an adaptive learning convolutional neural network (AL-CNN) for image classification, with results reported on a large-scale benchmark dataset that is widely accepted for performance evaluation. The AL-CNN architecture integrates convolutional, pooling, and fully connected layers. The model was systematically trained on a subset of the dataset and subsequently tested on an independent validation subset to evaluate its efficiency and generalization capability. In addition, optimization techniques such as data augmentation, dropout, and advanced activation functions were employed to further enhance model performance. The results, based on accuracy metrics, indicate the successful implementation of the proposed AL-CNN model for reliable and accurate image classification. This study demonstrates the potential of the AL-CNN approach to address various complexities in image classification, thereby enabling further innovation in this domain.

Corresponding Author:

Bharat Bhushan
School of Healthcare and Allied Sciences, GD Goenka University
Gurugram, India
Email: drbbhushan@gmail.com

1. INTRODUCTION

Image classification approaches used to be based on manual feature extraction methods complemented with classifiers such as support vector machines but happily the field has changed dramatically since the introduction of deep learning architectures, for example convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid networks that take advantage of hierarchical feature representation. Such networks allow for automatic learning of intricate patterns and image structural abstractions, which leads to performance improvement by a large margin. Very well-known networks, such as AlexNet, VGGNet, ResNet, LeNet, and DenseNet, have achieved incredibly high levels of classification accuracy [1]. Computer vision, as a result, has developed into one of the most exciting scientific disciplines of the 21st century with its applications ranging from healthcare diagnostics and autonomous vehicles to intelligent surveillance systems. Deep learning, a branch of artificial intelligence, makes use of compositions of artificial neurons to learn features automatically at multiple depths, which results in better understanding of visuals. Traditional rule-based techniques like edge detection, sliding window, and the Viola-Jones algorithm provide the basis for the enthusiast but have mainly been replaced by deep learning methods that are more adaptable and require fewer computations. CNNs, in particular, represent a major advancement due to their ability to learn directly from raw data without reliance on handcrafted features [2], [3]. Over time, CNN architectures have evolved, yielding models that range from highly deep to computationally efficient and generalisable forms [4]. AlexNet marked a significant breakthrough by introducing rectified linear unit (ReLU) activation, max-pooling, data augmentation, and dropout regularisation, achieving superior performance on large-scale datasets, albeit with increased computational cost and susceptibility to overfitting [5]. Conversely, LeNet-5 offered a lightweight design with efficient feature extraction but limited scalability for complex tasks [6]. The proposed adaptive learning convolutional neural network (AL-CNN) integrates the strengths of both models, combining low-level feature extraction with high-level semantic learning, thereby achieving an optimal balance between accuracy, generalisation, and computational efficiency.

Problem definition: building upon the success of convolutional neural networks in image classification, this paper introduces a novel method termed AL-CNN, which adapts learning parameters to accommodate feature complexity and variability across datasets, thereby improving classification performance. The proposed AL-CNN integrates adaptive feature extraction with optimized training algorithms to not only enhance classification accuracy but also improve generalization and robustness across diverse image datasets.

Objectives: to address the identified research gap and evaluate the effectiveness of the proposed hybrid architecture, the present study was undertaken with the following objectives:
- Develop an AL-CNN model incorporating an adaptive learning mechanism for effective feature extraction.
- Employ self-tuning hyperparameters to optimize the model with respect to data complexity.
- Integrate key architectural features from AlexNet and LeNet to improve classification accuracy.
- Conduct a comprehensive performance evaluation of the proposed AL-CNN model and provide directions for future research in the field.

Scope: this work focuses on classifying images from the Fashion-MNIST and CIFAR-10 datasets using the proposed AL-CNN architecture, implemented in Python. The scope of the study is constrained by the characteristics of the datasets and available computational resources. Future research directions may include scalability to larger datasets, real-time implementation, and integration with transformer-based models to further enhance performance.

Literature review: image classification techniques: the introduction of CNNs has marked a paradigm shift in the field of computer vision by enabling automated feature extraction and improved generalization performance. Subsequent advancements, such as ResNet in 2015 and DenseNet in 2017, have addressed fundamental challenges, including the vanishing gradient problem, while enhancing feature propagation and computational efficiency. More recently, vision transformers (ViT) have emerged as an alternative approach based on attention mechanisms for processing image patches effectively, thereby providing a complementary framework to traditional CNN-based methodologies for image analysis.

Ciresan *et al.* [7] highlighted the flexibility and performance of CNNs in image classification. The study proposed high-performance GPU-based variants of CNNs trained using online gradient descent. Experiments were conducted on the CIFAR-10 and MNIST datasets without relying on pre-wired or manually constructed feature extractors, demonstrating the adaptability of the proposed approach.

Nedeljkovic *et al.* [8] analyzed the application of fuzzy logic for image classification, emphasizing its effectiveness in generating naturalistic descriptions. However, approximately 30% of the samples were misclassified, primarily due to cloud dominance in the images. Furthermore, brighter regions often resulted in missed classifications, indicating the sensitivity of the method to such visual conditions.

Smart and Zhang [9] reviewed image recognition methods with a focus on multi-class classification tasks. The study examined dynamic classification techniques within genetic programming to address multi-class object classification problems and evaluated whether dynamic methods outperformed static approaches in handling complex classification scenarios.

Shan and Rodarmel [10] discussed hyperspectral imagery as a means to enhance the characterization of detailed object features. The study examined the benefits and effectiveness of principal component analysis (PCA) as a preprocessing technique for hyperspectral image analysis, highlighting its role in improving both classification accuracy and computational efficiency.

Latif *et al.* [11] focused on fruit classification and calorie estimation by presenting an extensive fruit image dataset. This dataset was specifically developed to support research in fruit classification and calorie estimation, demonstrating the critical importance of high-quality data in the development of robust deep learning models.

Maitlo *et al.* [12] introduced a new dataset designed for the inspection and classification of date fruits. The dataset includes multiple physical attributes and defect characteristics of date fruits, captured using advanced imaging technologies. In addition to enriching the available data resources for developing machine learning models in date fruit classification, this work also offers opportunities for achieving higher levels of automation in date fruit inspection and sorting processes.

AlexNet, introduced by Alex Krizhevsky in 2012 and later refined in 2017 [5], represented a pivotal advancement in the field of computer vision. The architecture comprises five convolutional layers followed by three fully connected layers, incorporating ReLU activation functions and dropout regularization, which significantly boosted interest in CNNs. VGGNet, developed by Oxford's Visual Geometry Group, emphasized increased network depth through architectures with 16–19 layers and small 3×3 convolutional filters, achieving high classification accuracy at the expense of increased computational complexity.

CNNs are a class of deep learning models widely employed for image classification and visual recognition due to their ability to automatically learn hierarchical spatial features directly from raw input data. A typical CNN architecture comprises an input layer, multiple hidden layers—including convolutional, pooling, and fully connected layers—and an output layer. Convolutional layers extract discriminative features from grid-structured data, while pooling layers reduce spatial dimensionality, improving computational efficiency and generalization. Fully connected layers integrate learned features using weighted linear combinations and nonlinear activations, and output layers employ functions such as SoftMax or sigmoid to generate class probabilities for classification tasks [13].

Network training involves feedforward propagation, loss computation using functions such as cross-entropy, and iterative weight optimization through backpropagation [13]. CNNs offer several advantages over traditional neural networks, including efficient parameter sharing, reduced risk of overfitting, and translation invariance, enabling robust detection of patterns regardless of their spatial location. Their hierarchical learning capability allows the automatic extraction of low-level features such as edges and textures in early layers and high-level semantic representations in deeper layers, significantly enhancing classification performance [14]. Convolution operations form the core of CNNs by generating feature maps through sliding filters, facilitating progressive feature abstraction while maintaining computational efficiency. Activation functions introduce non-linearity, with ReLU being the most widely adopted due to faster convergence, reduced vanishing gradient problems, and computational simplicity. While sigmoid and tanh activations are prone to saturation, sigmoid remains useful for binary classification outputs [15]. Pooling operations, including max pooling and average pooling, further enhance robustness by reducing feature dimensionality. Max pooling is particularly effective in preserving dominant features and improving invariance to minor spatial transformations. CNNs have found extensive applications across domains such as object detection, facial recognition, medical image analysis, satellite image classification, and automated disease diagnosis, underscoring their significance in modern computer vision research [16]-[20].

2. PROPOSED ALGORITHM

2.1. Adaptive learning convolutional neural network algorithm

Algorithm steps for image classification using a hybrid CNN (AlexNet+LeNet) on Fashion-MNIST as shown in Figure 1:

a. Initialize hyperparameters: set the following hyperparameters: dataset=`Fashion-MNIST` or `CIFAR-10`; batch size=128; learning rate=0.001; and number of epochs=10, 20, or 30.
b. Load and preprocess data: load the Fashion-MNIST dataset using Keras. Normalize pixel values to the range [0, 1] by dividing by 255.0. Reshape the training and testing data to include a single channel dimension for grayscale images.
c. Define the hybrid CNN model: create a Keras sequential model. Add an input layer of size (28×28×1), followed by convolutional layers with ReLU activation and varying filter sizes (32, 64, 128, and 256), max-pooling layers, a flatten layer, fully connected layers with 120 and 84 units using ReLU activation, and an output layer with a SoftMax classifier for 10 classes.
d. Compile the model: use the Adam optimizer with a learning rate of 0.001. Set the loss function to `sparse_categorical_crossentropy`, which is appropriate for integer-labeled multi-class classification. Track accuracy as the evaluation metric.
e. Train the model: train the model using the training dataset (`x_train`, `y_train`) with a batch size of 128

and the selected number of epochs. Use 10% of the training data for validation.

f. Evaluate the model: evaluate the trained model on the test dataset (`x_test`, `y_test`). Record the final test accuracy.
g. Extract validation loss and accuracy: retrieve and report the validation loss and validation accuracy
h. Make predictions: use the trained model to generate predictions for the test dataset. Convert prediction probabilities to class labels using the `np.argmax` function.
i. Generate the confusion matrix and summarize results: compute the confusion matrix using the true labels (`y_test`) and predicted labels (`y_pred`) with the `confusion_matrix()` function from `sklearn.metrics`. Visualize the confusion matrix using a heatmap (`sns.heatmap()`) with appropriate labels and a title. Report the dataset name, batch size, learning rate, number of epochs, final validation loss, final validation accuracy, and final test accuracy.

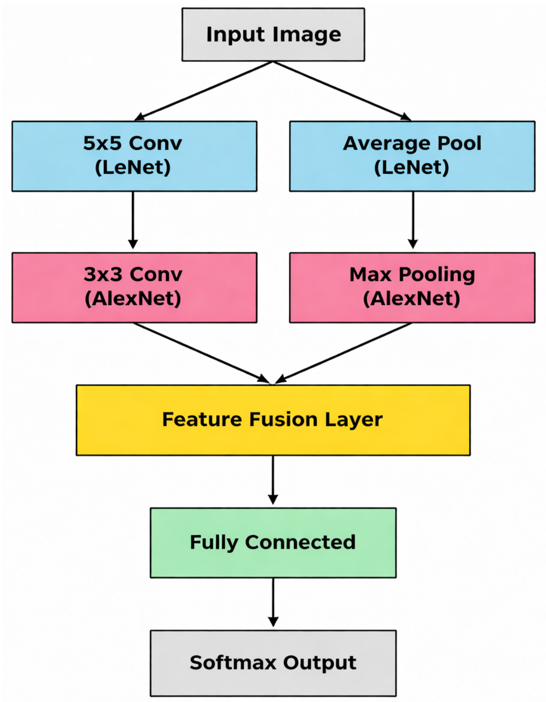

Figure 1. The steps of the proposed AL-CNN model structure

2.2. Comparative analysis of convolutional neural networks in image classification

CNN performance is commonly evaluated using benchmark datasets such as ImageNet, CIFAR-10/100, and MNIST, which provide standardized settings for assessing image classification capabilities across varying image resolutions and class complexities [21], [22]. CIFAR datasets are widely used for small-scale object recognition tasks, while ImageNet serves as a large-scale benchmark for deep CNN evaluation. Compared to traditional machine learning methods reliant on handcrafted features, CNNs automatically learn hierarchical representations, enabling superior generalization across diverse datasets, as reflected in the proposed AL-CNN architecture (Figure 2) [23]. Among prominent CNN architectures, VGGNet prioritizes accuracy at the cost of increased parameters, ResNet enables deeper networks through residual learning, and Inception captures multi-scale features efficiently using parallel convolutional paths [24]. These models highlight the inherent trade-off between classification accuracy and computational efficiency, often necessitating optimization strategies such as pruning and quantization [25]. The proposed hybrid AL-CNN addresses this trade-off by integrating efficient feature extraction with adaptive learning. Additionally, robustness to image variations and domain-specific performance remains critical, where data augmentation and improved generalization strategies further enhance CNN reliability across real-world applications [2].

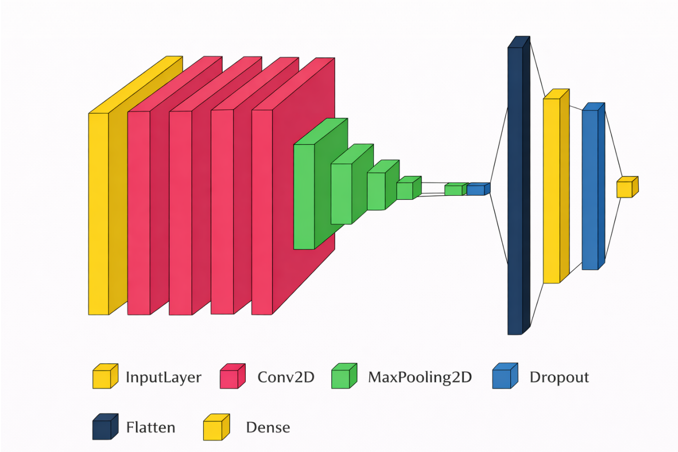

Figure 2. The schematic representation of the proposed AL-CNN model structure

3. METHOD

3.1. Data collection

Datasets used: this study utilises two benchmark datasets—Fashion-MNIST and CIFAR-10—to evaluate image classification performance. Fashion-MNIST comprises 60,000 training and 10,000 testing grayscale images of size 28×28 pixels across 10 clothing-related classes, serving as a more challenging alternative to the original MNIST dataset. CIFAR-10 consists of 60,000 colour images with a resolution of 32×32 pixels, divided into 10 object categories, with 50,000 images for training and 10,000 for testing. Compared to Fashion-MNIST, CIFAR-10 presents higher visual complexity and variability, making it suitable for assessing model generalisation across diverse image characteristics.

3.2. Data augmentation

To improve model generalization and mitigate overfitting, data augmentation techniques were incorporated during training. These techniques artificially expand the size of the training dataset through random transformations, enabling models to learn invariant and robust feature representations. For the Fashion-MNIST dataset, augmentation methods such as random rotation, shifting, and zooming were applied. For the more complex CIFAR-10 dataset, additional augmentation techniques, including random cropping, horizontal flipping, and color jittering, were employed. These augmentations increase data diversity and encourage the learning of generalized feature patterns, thereby improving overall classification accuracy.

3.3. Data preprocessing

Normalization, reshaping, and encoding: preprocessing is essential to ensure compatibility between the datasets and the machine learning models. Both the Fashion-MNIST and CIFAR-10 datasets were normalized by scaling pixel values to the range [0, 1], facilitating uniform training and faster convergence. The grayscale images in Fashion-MNIST were reshaped to include a single channel, whereas the CIFAR-10 images were preserved in their original RGB format. In addition, one-hot encoding was applied to the class labels, converting them into a binary matrix suitable for training deep learning models. These preprocessing steps ensured effective model compatibility and reliable training performance. For the CIFAR-10 dataset, pixel values of the color images were normalized by dividing by 255.

```
X_train = X_train.astype('float32')
X_test = X_test.astype('float32')
X_train /= 255
X_test /= 255
```

Experimental setup (AL-CNN): the experiments were implemented using Python with Keras and a TensorFlow backend on a GPU-enabled computing platform to enhance training speed and computational efficiency. The proposed AL-CNN model was evaluated on the Fashion-MNIST and CIFAR-10 benchmark datasets. Hyperparameters were initially selected based on established practices and further refined using adaptive learning strategies. Grid search was applied to optimise batch size, learning rate, and training epochs, while random search explored variations in dropout rates and pooling strategies. Model performance was assessed using validation accuracy to identify optimal configurations. NumPy and Scikit-learn libraries supported data processing and evaluation. The final training setup employed the Adam optimizer with a batch size of 128, using progressive training over 10, 20, and 30 epochs to improve model convergence and accuracy.

3.4. Model selection

The model architecture was selected based on the complexity and characteristics of each dataset. Since the Fashion-MNIST dataset comprises relatively simple grayscale images, a fully connected neural network (FCNN) was employed, as its dense layers are effective in capturing patterns and relationships among features for accurate clothing category classification.

For the CIFAR-10 dataset, a CNN was utilized due to its ability to extract spatial features. CNNs employ convolutional layers to learn hierarchical representations such as edges, textures, and object shapes, making them well suited for complex image datasets. Max-pooling layers were incorporated to reduce spatial dimensionality, lower computational complexity, and enhance translation invariance. By stacking multiple convolutional layers, the model was able to learn increasingly abstract and complex feature representations, thereby ensuring optimal performance for CIFAR-10 classification.

3.5. Training and evaluation

Model training was carried out on a Google Co-lab TPU environment using Python, Keras, and TensorFlow. The initial training was performed for 10 epochs with a batch size of 128, employing the Adam optimizer with a learning rate of 0.001. Hyperparameter tuning was conducted using a combination of grid search and random search to identify optimal training configurations. Grid search systematically explored different batch sizes (32, 64, and 128), learning rates (0.1, 0.01, 0.001, and 0.0001), and numbers of training epochs (10, 20, and 30). In contrast, random search enabled broader exploration of additional parameters, including dropout rates and activation functions. The Keras-Tuner library was utilized to automate the hyperparameter optimization process and to select the optimal combination based on validation accuracy.

Evaluation metrics: accuracy was selected as the primary performance metric, defined as the percentage of correctly classified instances. Categorical cross-entropy loss was employed as the loss function, as the task involved multi-class classification. Accuracy provided an overall assessment of model performance, while cross-entropy loss was used to evaluate potential overfitting and underfitting. Together, these metrics enabled a comprehensive comparison of model performance across the Fashion-MNIST and CIFAR-10 datasets.

4. RESULTS AND DISCUSSION

The performance of both the initial and refined models was evaluated using plots illustrating accuracy and loss trends over the training epochs. In the case of the FCNN, the model was trained for 30 epochs and demonstrated suitable performance on the Fashion-MNIST dataset. As shown in Figure 3, the accuracy trend reflects the learning behavior of the FCNN trained on Fashion-MNIST over 30 epochs.

```
Epoch 21/30
469/469 ━━━━━━━━━━━━━━━━━━━━ 57s 122ms/step - accuracy: 0.9775 - loss: 0.0620 - val_accuracy: 0.9029 - val_loss: 0.3937
Epoch 22/30
469/469 ━━━━━━━━━━━━━━━━━━━━ 58s 124ms/step - accuracy: 0.9771 - loss: 0.0617 - val_accuracy: 0.9110 - val_loss: 0.3969
Epoch 23/30
469/469 ━━━━━━━━━━━━━━━━━━━━ 78s 115ms/step - accuracy: 0.9817 - loss: 0.0502 - val_accuracy: 0.9101 - val_loss: 0.3637
Epoch 24/30
469/469 ━━━━━━━━━━━━━━━━━━━━ 54s 115ms/step - accuracy: 0.9828 - loss: 0.0471 - val_accuracy: 0.9110 - val_loss: 0.4050
Epoch 25/30
469/469 ━━━━━━━━━━━━━━━━━━━━ 56s 120ms/step - accuracy: 0.9815 - loss: 0.0478 - val_accuracy: 0.9073 - val_loss: 0.4116
Epoch 26/30
469/469 ━━━━━━━━━━━━━━━━━━━━ 54s 115ms/step - accuracy: 0.9845 - loss: 0.0410 - val_accuracy: 0.9109 - val_loss: 0.4390
Epoch 27/30
469/469 ━━━━━━━━━━━━━━━━━━━━ 84s 119ms/step - accuracy: 0.9880 - loss: 0.0318 - val_accuracy: 0.9117 - val_loss: 0.4250
Epoch 28/30
469/469 ━━━━━━━━━━━━━━━━━━━━ 81s 117ms/step - accuracy: 0.9877 - loss: 0.0363 - val_accuracy: 0.9109 - val_loss: 0.4471
Epoch 29/30
469/469 ━━━━━━━━━━━━━━━━━━━━ 55s 117ms/step - accuracy: 0.9863 - loss: 0.0368 - val_accuracy: 0.9107 - val_loss: 0.4609
Epoch 30/30
469/469 ━━━━━━━━━━━━━━━━━━━━ 55s 118ms/step - accuracy: 0.9877 - loss: 0.0325 - val_accuracy: 0.9018 - val_loss: 0.4885
Test Loss: 0.4885
Test Accuracy: 0.9018
```

Figure 3. Output of set trained for 30 epochs, Fashion-MNIST

Figure 4 presents the confusion matrix of the AL-CNN model trained on Fashion-MNIST, illustrating class-wise prediction performance and misclassification patterns. Figure 5 depicts training and validation accuracy and loss curves for the Fashion-MNIST dataset over 30 training epochs. The performance of AL-CNN on CIFAR-10 was further analyzed using classification outputs, confusion matrices, and training curves. Figure 6 shows sample classification results of the proposed AL-CNN on the CIFAR-10 dataset after 30 training epochs. Figure 7 presents the confusion matrix for CIFAR-10, highlighting correct and incorrect class predictions. Figure 8 illustrates the accuracy and loss behavior of the proposed AL-CNN model on CIFAR 10 dataset.

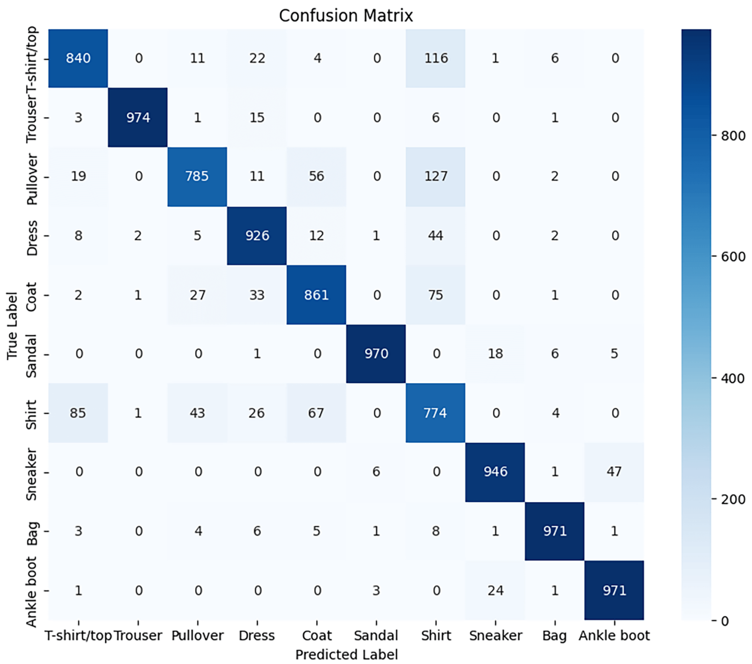

Figure 4. Confusion matrix of set trained for 30 epochs, Fashion-MNIST

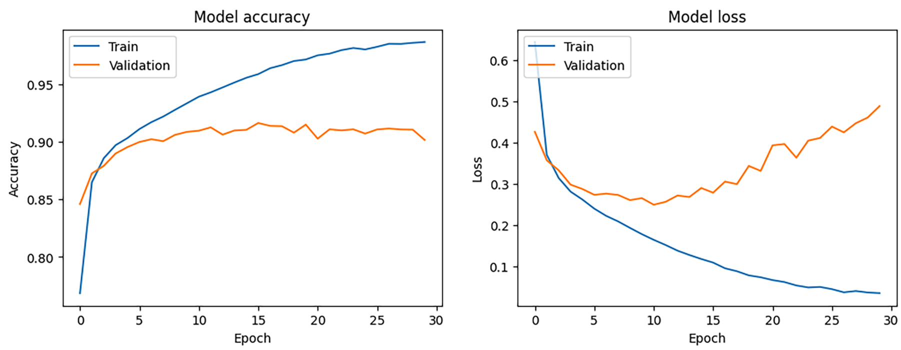

Figure 5. Graphical representation of model accuracy and model loss of set trained for 30 epochs, used on fashion MNIST

Epoch 1/30: accuracy: 0.4168 - loss: 1.6012 - val_accuracy: 0.5349 - val_loss: 1.2789

Epoch 2/30: accuracy: 0.5635 - loss: 1.2248 - val_accuracy: 0.6391 - val_loss: 1.0115

...

Epoch 30/30: accuracy: 0.9126 - loss: 0.2458 - val_accuracy: 0.7835 - val_loss: 0.7547

Figure 6. Output of set trained for 30 epochs, used on CIFAR-10

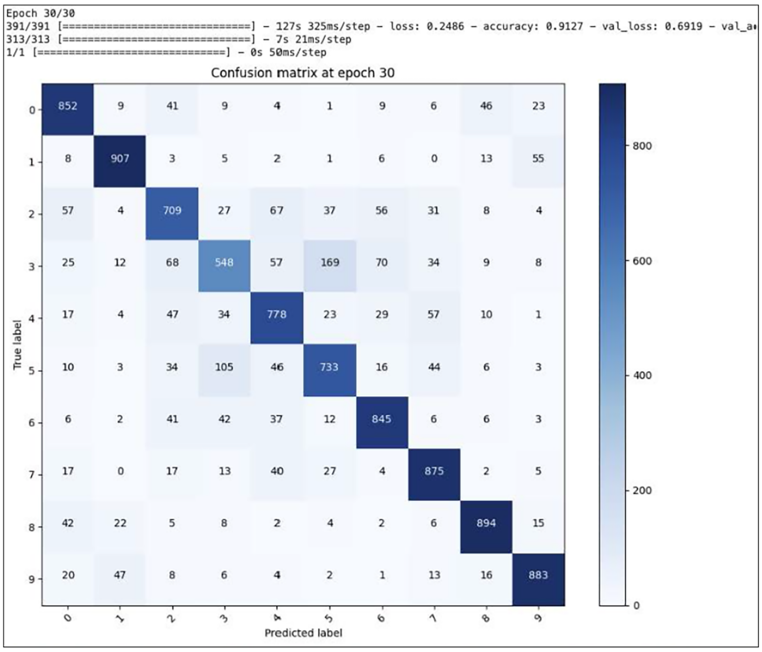

Figure 7. Confusion matrix of set trained for 30 epochs, used on CIFAR-10

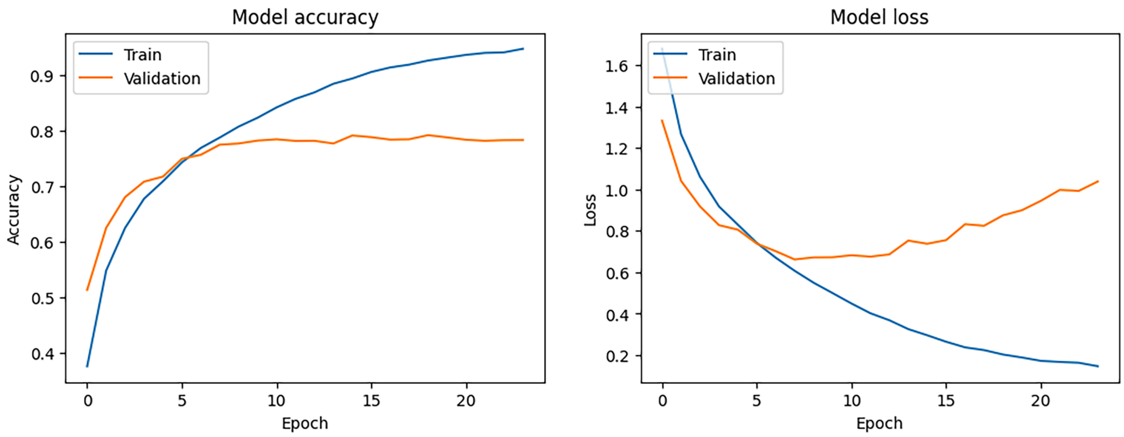

Figure 8. Graphical representation of model accuracy and model loss of set trained for 30 epochs, used on CIFAR-10

These results highlight the versatility of different neural network architectures when applied to datasets with varying levels of complexity and underscore the impact of advanced models such as AlexNet and LeNet in improving classification accuracy. Accuracy and loss curves for both training and validation phases were analyzed to visualize the learning behavior of the models and to monitor effective training as well as potential overfitting. Table 1 presents a summarized comparison of the accuracy and loss metrics for different models trained on the Fashion-MNIST and CIFAR-10 datasets. The table reports key parameters, including batch size, learning rate, number of epochs, validation loss, validation accuracy, and final test accuracy. This comparative analysis clearly demonstrates the performance improvement achieved by the proposed AL-CNN architecture over the baseline models.

4.1. Analysis of result

The performance of the proposed models was evaluated through comparison with baseline models and existing methods to identify performance improvements and limitations. The baseline model, a basic

CNN, achieved an accuracy of 83% at the 20th epoch. In contrast, the AL-CNN model, which incorporates a more sophisticated architecture and optimized hyperparameter settings, demonstrated substantial performance improvement, achieving an accuracy exceeding 91% at the 30th epoch.

Table 1. Performance metrics of different models and datasets

Dataset	B_S	L_R	N_E	V_L	V_A (%)	F_A (%)
Fashion-MNIST with fully connected layers (CNN)	128	0.001	10	0.2717	90.10	89.67
Fashion-MNIST with CNN	128	0.001	20	0.2601	90.78	90.42
Fashion-MNIST with AL-CNN (30 epochs)	128	0.001	30	0.3935	91.28	91.01
CIFAR-10 with CNN (20 epochs)	128	0.001	20	0.8358	72.15	71.00
CIFAR-10 with AL-CNN (20 epochs)	128	0.001	20	0.6540	79.28	76.78
CIFAR-10 with AL-CNN (30 epochs)	128	0.001	30	0.7547	78.35	91.26

When compared with existing methods, the AL-CNN model exhibited performance comparable to state-of-the-art results on both the Fashion-MNIST and CIFAR-10 benchmark datasets. This comparison highlights the effectiveness of the proposed model, demonstrating improved performance over less complex baseline models while also indicating areas that may benefit from further optimization and refinement.

Confusion matrix analysis: confusion matrices were employed to perform a detailed analysis of model performance. These matrices are useful for identifying specific misclassifications and understanding class-wise error patterns where the model fails. For instance, the confusion matrix of the AL-CNN model illustrates classification accuracy for each class and highlights frequently misclassified categories, which can be addressed through targeted model tuning. Analyzing these matrices enables informed adjustments to further enhance predictive accuracy and overall model performance.

5. CONCLUSION

This research investigated the potential of convolutional neural networks for image classification using the CIFAR-10 dataset. Experimental results demonstrate that advanced architectures and training strategies, including deeper networks, regularization techniques, and data augmentation, significantly improve classification performance. The AL-CNN model achieved an accuracy exceeding 91%, underscoring its effectiveness in addressing challenging image classification problems. These findings highlight the proposed model as a robust and powerful approach with broad applicability across domains such as surveillance, manufacturing, and agriculture.

Future work: future research may extend this work by evaluating the proposed approach on larger and more diverse datasets, such as ImageNet, which is a large-scale image ontology structured. Additional performance improvements may be achieved by exploring advanced architectures, including ViTs and hybrid models. Furthermore, transfer learning techniques could be employed to enhance accuracy and reduce training time by fine-tuning pre-trained models for tasks with limited data. Self-supervised learning approaches that leverage unlabeled data may further improve model robustness and generalization. Finally, enhancing model interpretability will be critical for understanding AL-CNN decision-making processes and improving user trust in real-world applications.

FUNDING INFORMATION

This research did not receive specific grants from funding agencies in the public, commercial or non-profit sectors. The study was conducted and managed entirely by the authors.

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
Monika Chawla	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	✓
Rashmi Agrawal	✓	✓		✓	✓	✓		✓		✓	✓	✓		
Bharat Bhushan	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓		✓	✓

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

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Authors state no conflict of interest.

DATA AVAILABILITY

The authors confirm that the data supporting the findings of this study are available in the article, but if more information is needed on the research. Data supporting the findings of this study are available upon request from the corresponding author [initials: BB]. upon reasonable request.

REFERENCES

[1] M. Swapna, D. Y. K. Sharma, and D. B. Prasad, "CNN Architectures: Alex Net, Le Net, VGG, Google Net, Res Net," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 6, pp. 953–959, 2020, doi: 10.35940/ijrte.f9532.038620.

[2] J. Plested, M. Phiri, and T. Gedeon, "Deep transfer learning for image classification: a survey," *Artificial Intelligence Review*, vol. 59, no. 3, 2026, doi: 10.1007/s10462-026-11491-z.

[3] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," *Computer Science*, 2009.

[4] M. M. Taye, "Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions," *Computation*, vol. 11, no. 3, pp. 1–23, Mar. 2023, doi: 10.3390/computation11030052.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[6] W. Cui, Q. Lu, A. M. Qureshi, W. Li, and K. Wu, "An adaptive LeNet-5 model for anomaly detection," *Information Security Journal*, vol. 30, no. 1, pp. 19–29, Jan. 2021, doi: 10.1080/19393555.2020.1797248.

[7] D. C. Cireşan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Flexible, high performance convolutional neural networks for image classification," in *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, 2011, pp. 1237–1242, doi: 10.5591/978-1-57735-516-8/IJCAI11-210.

[8] I. Nedeljkovic, C. Vi, and W. G. Vi, "Image Classification Based on Fuzzy Logic," in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2004, pp. 3–7.

[9] W. Smart and M. Zhang, "Classification strategies for image classification in genetic programming," in *Proceeding of Image and Vision Computing Conference*, Palmerston North, New Zealand, 2003, pp. 402–407.

[10] J. Shan and C. Rodarmel, "Principal Component Analysis for Hyperspectral Image Classification," *Surveying and Land Information Science*, vol. 62, no. 2, pp. 115–123, 2002.

[11] G. Latif, N. Mohammad, and J. Alghazo, "DeepFruit: A dataset of fruit images for fruit classification and calories calculation," *Data in Brief*, vol. 50, pp. 1–6, 2023, doi: 10.1016/j.dib.2023.109524.

[12] A. K. Maitlo, R. A. Shaikh, and R. H. Arain, "A novel dataset of date fruit for inspection and classification," *Data in Brief*, vol. 52, pp. 1–7, Feb. 2024, doi: 10.1016/j.dib.2023.110026.

[13] G. Cai, "Advanced Image Classification Using Convolutional Neural Networks," *Science and Technology of Engineering, Chemistry and Environmental Protection*, vol. 1, no. 9, Oct. 2024, doi: 10.61173/117trk02.

[14] X. Liu, L. Song, S. Liu, and Y. Zhang, "A review of deep-learning-based medical image segmentation methods," *Sustainability*, vol. 13, no. 3, pp. 1–29, Jan. 2021, doi: 10.3390/su13031224.

[15] I. T. Ahmed, C. S. Der, N. Jamil, and M. A. Mohamed, "Improve of contrast-distorted image quality assessment based on convolutional neural networks," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 6, pp. 5604–5614, Dec. 2019, doi: 10.11591/ijece.v9i6.pp5604-5614.

[16] W. Ding *et al.*, "FTransCNN: Fusing Transformer and a CNN based on fuzzy logic for uncertain medical image segmentation," *Information Fusion*, vol. 99, Nov. 2023, doi: 10.1016/j.inffus.2023.101880.

[17] L. Li, Q. Liu, X. Shi, Y. Wei, H. Li, and H. Xiao, "UCFilTransNet: Cross-Filtering Transformer-based network for CT image segmentation," *Expert Systems with Applications*, vol. 238, Mar. 2024, doi: 10.1016/j.eswa.2023.121717.

[18] J. Chen *et al.*, "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation," *arXiv preprint*, 2021, doi: 10.48550/arXiv.2102.04306.

[19] G. Andrade-Miranda, V. Jaouen, O. Tankyevych, C. C. L. Rest, D. Visvikis, and P.-H. Conze, "Multi-modal medical Transformers: A meta-analysis for medical image segmentation in oncology," *Computerized Medical Imaging and Graphics*, vol. 110, Dec. 2023, doi: 10.1016/j.compmedimag.2023.102308.

[20] I. Shiri *et al.*, "Multi-institutional PET/CT image segmentation using federated deep transformer learning," *Computer Methods and Programs in Biomedicine*, vol. 240, pp. 1–16, Oct. 2023, doi: 10.1016/j.cmpb.2023.107706.

[21] L. Karacan, "Multi-image transformer for multi-focus image fusion," *Signal Processing: Image Communication*, vol. 119, Nov. 2023, doi: 10.1016/j.image.2023.117058.

[22] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms," *arXiv preprint*, 2017, doi: 10.48550/arXiv.1708.07747.
[23] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA: IEEE, Jun. 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.
[24] Y. Zhao *et al.*, "A transformer-based hierarchical registration framework for multimodality deformable image registration," *Computerized Medical Imaging and Graphics*, vol. 108, Sep. 2023, doi: 10.1016/j.compmedimag.2023.102286.
[25] H. T. Mustafa, P. Shamsolmoali, and I. H. Lee, "TGF: Multiscale transformer graph attention network for multi-sensor image fusion," *Expert Systems with Applications*, vol. 238, Mar. 2024, doi: 10.1016/j.eswa.2023.121789.

BIOGRAPHIES OF AUTHORS

Monika Chawla completed her M.E. in Computer Science from LIMAT, MDU Rohtak, India, in 2007 and is pursuing a Ph.D. in Computer Science and Engineering at Manav Rachna Institute of Research and Studies, Faridabad. She is an Assistant Professor at the School of Computer Applications, Banarsidas Chandiwala Institute of Information Technology, New Delhi, with over 20 years of academic experience. Her research interests include machine learning, deep learning, image processing, artificial intelligence, and optimization, with publications in reputed journals. She can be contacted at email: monika.chawla12@gmail.com.

Dr. Rashmi Agrawal is a Professor in the Department of Computer Applications at Manav Rachna International Institute of Research and Studies, Faridabad, India. She is Ph.D. and UGC-NET qualified, with more than 20 years of experience in teaching and research. She is a Life Member of the Computer Society of India and a Senior Member of IEEE. She serves as a book series editor with CRC Press (Taylor & Francis) and Wiley, has published extensively in reputed indexed journals and conferences, edited books with international publishers, supervises Ph.D. scholars in emerging research areas, and serves as an Associate Editor for Elsevier journals. She can be contacted at email: rashmi.sca@mriu.edu.in.

Bharat Bhushan is a Professor in the School of Healthcare and Allied Sciences at GD Goenka University, Gurugram, India. He previously served as Dean (Physiotherapy) and Associate Professor at Baba Mastnath University, Rohtak, with over 18 years of academic service, including seven years as Dean. He holds an M.Tech. in Software Engineering from UIET, MDU Rohtak, and has more than 23 years of teaching experience. His research interests focus on image processing, particularly human postural analysis, with contributions to international projects (RISHII) and scholarly publications. He can be contacted at email: drbbhushan@gmail.com.