International Journal of Intelligent Data and Machine Learning (IJIDML)
eISSN: 3087-4262
Volume03 Issue07 | Page No: - 31-39

Received: 30 May 2026, Revised: 15 June 2026, Accepted: 28 June 2026, Published: 24 July 2026

DOI: https://doi.org/10.55640/ijidml-v03i07-04

An Efficient Deep Learning Framework Model for High-Accuracy Image Visual Classification

Mr. Mohit Sahu [1]

[1] Assistant Professor, Department of Computer Sciences and Applications, Mandsaur University, Mandsaur, India

ABSTRACT: The term "image" refers to a two-dimensional array of pixels used to depict an item, scene, or pattern in the visual domain. Digital imaging involves the use of imaging equipment, such as cameras, to either capture or fabricate pictures from light or electromagnetic radiation. Training algorithms to identify and classify photos by content is the essence of image classification, a machine learning process. This is accomplished by utilizing the CIFAR-10 dataset, which comprises sixty thousand colour photographs categorized into ten distinct item types, and a framework for high-accuracy picture visual categorization that is based on Recurrent Neural Networks (RNNs). Image normalization, one-hot label encoding, and data augmentation techniques are used to preprocess the dataset in order to prevent overfitting and increase model generalization. In order to acquire useful feature representations and execute reliable image classification, the suggested RNN design has a thick output layer after two recurrent layers. A 97.2% accuracy (ACC) rate, a 97.0% precision (PRE) rate, a 97.6% recall (REC) rate, and an F1-score (F1) of 97.4% were all achieved by the suggested model in the experiments. The suggested RNN achieves a higher classification accuracy (95.49%) than CNN (75.7%), ResNet50 (94.1%), and MobileNetV3 (95.49%), according to a comparison with current deep learning models. These results show that the suggested framework is a high-performance, dependable, and effective way to classify objects in images.

KEYWORDS: Deep learning, Image classification, Convolutional neural networks, Automated recognition, Image Visual Classification.

1. Introduction

Smart and effective picture visual categorization systems are in high demand due to the exponential growth of digital picture data generated by many sources such as cameras, sensors, satellites, medical equipment, and multimedia platforms [1][2]. Image classification aims to automatically classify photos into predetermined groups according to their visual attributes, object features, and learned feature representations. This is one of the most critical tasks in computer vision [3]. Effective image classification enables machines to understand and interpret visual information, supporting various applications such as medical image diagnosis, agricultural monitoring, autonomous vehicles, biometric identification, surveillance systems, remote sensing, and industrial quality inspection [4][5][6]. The increasing availability of large-scale image datasets has further accelerated research in automated visual recognition systems, requiring advanced models capable of handling complex variations in image content, including changes in illumination, background, object orientation, scale, and visual similarity among different categories [7][8].

Traditional ML-based image classification methods primarily rely on manually designed feature extraction techniques, such as texture, shape, and colour descriptors, followed by classification algorithms for decision-making [9][10]. However, these handcrafted approaches require expert knowledge, involve complex feature engineering processes, and often fail to effectively represent high-level semantic information from complex images. As image datasets continue to increase in size and diversity, conventional methods face limitations in terms of adaptability, scalability, and classification ACC [11].

Neural networks can now automatically learn hierarchical characteristics from raw photos, thanks to recent advancements in DL, which has revolutionized image categorization [12]. Automated hierarchical visual representations may be found by DL models by way of lower-level learning of fundamental characteristics like textures, patterns, and edges, and by way of deeper-level capturing of sophisticated object-level information [13].

The ACC, robustness, and generalization have been enhanced in comparison to classic ML methods, thanks to these features. Since CNNs are so good at extracting spatial characteristics via convolution operations, they have largely replaced all other image classification architectures [14]. However, for CNN-based models to work optimally, a lot of computing resources, parameter adjustment, and training data on a large scale are usually required [15][16][17]. RNNs, which are mainly developed for sequential data processing, provide an alternative learning mechanism by maintaining hidden states to capture dependencies among sequential inputs. Although images are naturally represented as two-dimensional spatial data, they can be transformed into sequential feature representations, allowing RNNs to learn relationships among image components. One reason RNNs are used for picture visual categorization is their capacity to grasp temporal-like patterns and sequential relationships. By utilizing recurrent learning mechanisms, RNN-based models provide an alternative approach for extracting meaningful visual representations and improving classification performance while exploring beyond conventional convolution-based methods.

A. Motivation and Contribution

A visual classification framework for photos utilizing RNNs capable of accurately identifying various item categories and generating sound generalizations is the target of this research. Although existing DL models have achieved significant success, challenges remain in designing efficient architectures that can effectively learn meaningful visual representations with reduced complexity. Therefore, this research proposes an RNN-based classification approach that utilizes sequential feature learning for image recognition tasks. This study's key findings are mentioned below:

- An RNN-based framework is proposed for image visual classification using sequential feature representation learning.
- The model is trained and tested using the CIFAR-10 dataset, which contains sixty thousand RGB photographs of objects belonging to ten distinct categories.
- Comprehensive preprocessing techniques, including normalization, resizing, data augmentation, and one-hot encoding, are applied to improve model robustness and classification performance.
- An efficient RNN architecture consisting of recurrent layers and a dense output layer is developed for multi-class image classification.
- The ACC, PRE, REC, and F1 are some of the performance indicators used to assess the suggested model's utility in classification tasks.
- The suggested method is compared to well-known DL models like CNN, ResNet50, and MobileNetV3 to show how well it works for classifying images visually.

B. Justification and Novelty

The goal of this research is to improve feature learning and classification performance using a dependable and effective visual classification framework for images that can properly identify a variety of item types. An innovative approach was taken in designing an RNN-based architecture that incorporates efficient preprocessing and data augmentation techniques. This was done to improve the model's resilience and ability to generalize. Unlike conventional image classification approaches, the proposed framework efficiently learns discriminative visual features and provides a dependable solution for image visual classification across diverse applications.

C. Organization of the Paper

The paper is structured in the following way: Research in this area is detailed in Section II. Section III of the study details the procedures and methods used to gather data. The analysis summarized in Section IV, and the paper's conclusion presented in Section V.

2. Literature Review

The increasing need for fast architectures that can classify visual images has put a spotlight on the creation of lightweight DL models. To strike a balance between PRE and efficiency, many lightweight designs have been suggested. For example, S et al., (2026) used transfer learning and a lightweight convolutional neural network architecture dubbed EfficientNet-B0, CT scans may be classified as either normal, cyst, stone, or tumour. Criteria for evaluation were F1, REC, ACC, and PRE. With an overall ACC of 84.02%, the model demonstrated great performance for the cyst and tumour classes using a publicly accessible CT Kidney dataset. It also highlighted the significance of class imbalance and visual likeness as hurdles in stone categorization [18]. Another worked by K, S and Gayathri, (2026) provides a method for the real-time categorization of manufacturing problems using an algorithmic approach based on ML and image processing. The three supervised ML models that get these features are SVM, RF, and KNN. A curated dataset was used for experimental assessments, and the RF classifier was shown to have the best ACC (94.1%), stability, and performance of the bunch [19].

Then, Saadati-Monem et al. (2025) The goal of learning-based image compression methods is to improve compression performance by allowing for more flexible feature space for computer vision tasks. This makes it easy to mix these methods with analytical tasks. The main objective of this project is to train a network that can classify images and a network that can compress them using learning-based techniques in an interactive way. The findings show that this strategy stabilizes the training process, increases ACC for picture classification using the extracted features, and retains and enhances the compression network's performance compared to standard methods. The suggested approach improves performance by 7.99% on average [20].

Yuan, Xiong and Chen (2024) aim to use machine vision to look at the phenotypes and classify the varieties of waxy corn seeds, tackling the problems that come up when varieties mix and slow down industry growth. A variety of ML techniques, each with its own set of advantages, were used to successfully categorize waxy corn seeds. These algorithms included K-Nearest Neighbours (KNN), Linear Discriminant Analysis (LDA), Decision Trees (DT), and Support Vector Machines (SVM). The following settings were used by the five models to attain high ACC rates, which exceeded 90%: DT at 0.922 and KNN at 0.947 [21]. Then, Krishna Prasad, Amudha and R, (2023) suggested comparing SAR image classification using a supported vector machine with a new greedy hierarchical learning approach. Twenty samples were selected from the Kaggle learning platform's dataset. In the MATLAB simulation, support vector machine attained a classification ACC of 80.41%, whereas innovative greedy hierarchical learning reached 93.08%. There is a significant result of 0.009 (p 0.05) according to the SPSS analysis. On the provided synthetic aperture radar image classification dataset, the new greedy hierarchical learning algorithm considerably beats the support vector machine in terms of ACC [22].

Lastly, Yang et al. (2022) assume that learning via a lightweight network increases computation efficiency and that collecting rich feature representation would help tyre classification. Afterwards, they suggest a two-step training process that is user-friendly and effective: 1) They project pictures to latent space using rich feature representation, and then use a variational auto-encoder framework that is limited by contrastive learning to create a feature extractor. 2) The characteristics collected by the previously learned encoder are used to train a one-layer linear classification network. On the tyre pattern dataset, the top-1 ACC is 89.8% and the top-5 ACC is 96.6%, proving that the technique is successful [23]. Table I summarizes current studies on image visual classification, including the models suggested, datasets used, key results, and difficulties encountered.

TABLE I. RELATED WORK SUMMARY ON IMAGE VISUAL CLASSIFICATION USING MACHINE LEARNING TECHNIQUES

Author	Proposed Work	Dataset	Results	Limitations & Future Work
S et al. (2026)	Proposed an EfficientNet-B0 transfer learning framework for multi-class kidney CT image classification.	Public CT Kidney dataset	Achieved 84.02% overall accuracy with strong performance for cyst and tumour classification.	Performance was affected by class imbalance and visual similarity in stone images. Future work should improve minority class recognition and model robustness.
K, S. and Gayathri (2026)	Developed a real-time manufacturing defect classification system using image processing with SVM, RF, and KNN classifiers.	Manufacturing defect image dataset	Random Forest achieved the highest accuracy of 94.1%.	Future work should evaluate larger industrial datasets and incorporate deep learning models for improved performance.
Saadati-Monem et al. (2025)	Proposed an interactive learning-based image compression network integrated with image classification.	Standard image compression and image classification datasets	Achieved an average 7.99% improvement in classification performance while maintaining compression quality.	Future work should validate the approach on larger datasets and real-time computer vision applications.
Yuan, Xiong and Chen (2024)	Proposed a machine vision framework for waxy corn seed classification using phenotypic feature extraction with KNN, LDA, DT, and SVM.	Waxy corn seed image dataset (three varieties)	All models achieved over 90% accuracy, with KNN 94.7% and DT 92.2%.	The study considered only three seed varieties. Future work should include more crop varieties, larger datasets, and deep learning approaches.
Krishna Prasad, Amudha and R. (2023)	Compared Greedy Hierarchical Learning with SVM for SAR image classification.	Kaggle SAR image dataset (20 samples)	Greedy Hierarchical Learning achieved 93.08% accuracy, outperforming SVM (80.41%).	The dataset size was very small. Future work should evaluate larger SAR datasets and advanced deep learning models.

Yang et al. (2022)	Proposed a lightweight two-stage framework using Variational Autoencoder and contrastive learning for tyre image classification.	Tire pattern image dataset	Achieved 89.8% Top-1 accuracy and 96.6% Top-5 accuracy.	Future work should validate the framework on larger and more diverse image datasets.

Research gaps: Existing image classification approaches often rely on handcrafted features, transfer learning, or application-specific models that may not generalize effectively across diverse image categories. Several studies also report limitations related to class imbalance, small datasets, and insufficient feature representation, which reduce classification robustness. Furthermore, limited attention has been given to leveraging recurrent neural networks for extracting discriminative visual features in general-purpose image classification tasks. Therefore, there is a need for an efficient RNN-based framework with effective preprocessing and feature learning to improve classification ACC and generalization.

3. Methodology

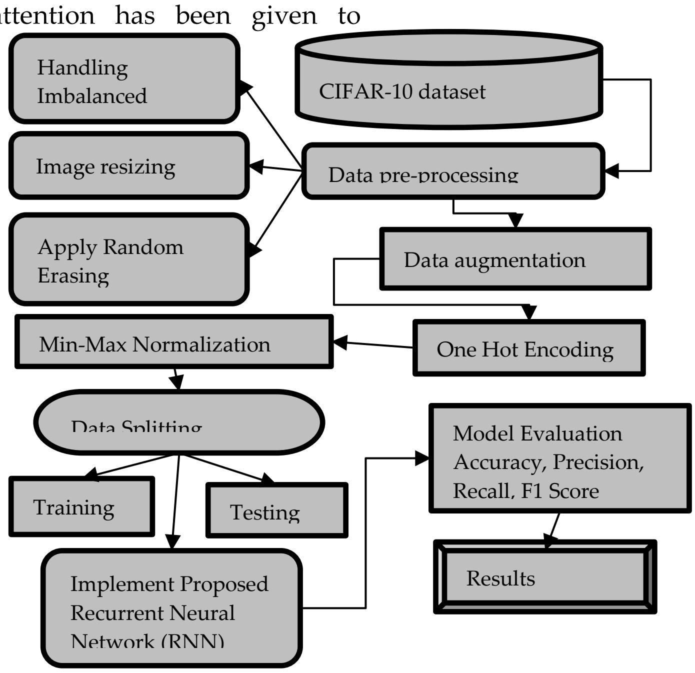

Fig. 1. Proposed flowchart for Image Classification using DL model

The section that follows provides a detailed explanation of the intended method:

The proposed methodology is developed for image visual classification using the CIFAR-10 dataset. Image normalization, resizing, min-max scaling, data augmentation, and one-hot encoding of class labels are some of the preprocessing procedures that are applied to the dataset in order to improve its quality and the efficiency of its model learning. Using stratified sampling, the preprocessed dataset is divided into two parts: one for training and another for testing. The former receives 80% of the data, while the latter receives 20%. This maintains a balanced distribution of classes. Afterwards, a multi-class classification model is constructed using an RNN, which consists of two recurrent layers and a fully connected dense output layer. This model is able to extract sequential feature representations from the input pictures. Using suitable hyperparameters, the model is trained and improved. ACC, PRE, REC, F1, loss analysis, and confusion matrix visualization are some of the performance indicators used to evaluate its efficacy. The evaluation measures offered here thoroughly test the model's generalizability and classification abilities. The suggested RNN model-based image visual classification procedure is shown in Figure. 1.

A. Data Collection and Preprocessing

DL and ML researchers use the CIFAR-10 dataset as a standard for picture classification tasks. There are a total of sixty thousand colour photos, each measuring thirty-two by thirty-two pixels, which are organized into ten different item classes: aircraft, vehicle, animal, ship, truck, frog, horse, tree, and deer. With 6,000 photos in each class and 50,000 in training and 10,000 in testing, the dataset is usually rather large. Due to the uniform distribution of photos across all classes, this dataset is ideal for testing classification methods. Researchers may evaluate the performance of various DL architectures under uniform settings on CIFAR-10 because of its balanced structure, modest size, and diverse object categories. Developing, training, and evaluating image classification algorithms make extensive use of the dataset.

Figure 2 displays samples from the CIFAR-10 dataset, with a single picture representing each of the ten item types: aircraft, car, bird, deer, dog, frog, horse, cart, ship, and truck. These examples showcase several classes and provide a high-level overview of the dataset used to train and evaluate the image classification algorithm.

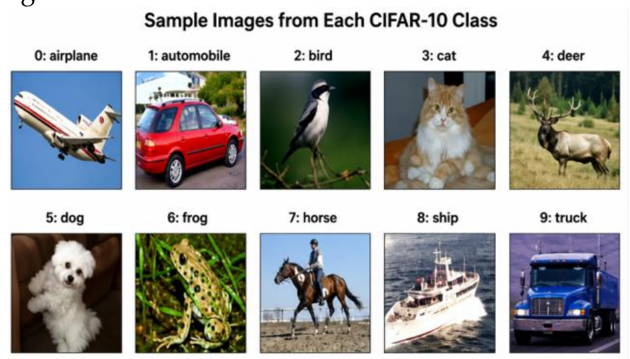

Fig. 2. Sample images

Fig. 3. Training Set Class Distribution

There is a perfect distribution of training pictures among the ten classes in the CIFAR-10 dataset, as seen in the Training Set Class Distribution (Figure 3). This balanced distribution helps to reduce class bias when training the model and ensures that the image classification results are fair and reliable.

B. Image Normalization

Using the formula Pixel Value = Pixel Value / 255.0, normalize all images by reducing the intensity values of the pixels from their original 0-255 range to 0-1. This normalization improves numerical stability, accelerates model convergence, and reduces the effect of varying pixel intensities during training.

C. Label Encoding

A technique that facilitates multi-class classification is the transformation of class labels into one-hot encoded vectors. Since the dataset contains ten distinct classes, and may conceive of each label as a 10-dimensional binary vector, where 1 denotes the correct class and 0 denotes the unsuitable ones.

D. Data Augmentation

The training dataset is the only one that uses data augmentation to make the model more generalizable and less prone to overfitting. The following are the parameters for the fixed augmentation: horizontal flip (with a probability of 0.5), rotation (with a margin of ±15°), width shift (10%), height shift (10%), and zoom (10%). The original class labels are preserved while these changes provide various training examples.

E. Dataset Splitting

The pre-processed dataset is stratified-sampled so that each of the 10 classes has an equal chance of being in the training or testing subset. Nearly 4,800 training pictures and 1,200 testing images are produced for each class, for a total of 48,000 training images and 12,000 testing images.

F. Proposed Recurrent Neural Network (RNN) Model

The proposed RNN model is developed to perform image visual classification using the CIFAR-10 dataset by learning sequential feature representations. Each 32 x 32 × 3 RGB image is transformed into 32 different time steps, with 96 features (32 × 3 each time step), because RNNs analyze data sequentially. By storing data from earlier time steps in its hidden state, RNNs are able to learn about input feature relationships. At time step t, the hidden state is calculated using Equation. (1):

$$h_t = \phi(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \quad (1)$$

The variables x_t represent the input at time step t, h_{t-1} stands for the previous hidden state, W_{xh} is the input-to-hidden weight matrix, W_{hh} is the recurrent hidden-to-hidden weight matrix, b_h is the bias vector, and $\phi(\cdot)$ is the ReLU activation function.

A fully linked output layer follows two SimpleRNN layers in the suggested design. The SimpleRNN architecture begins with a 128-hidden-unit layer activated using ReLU and set to return_sequences=True. The following is a 30-dropout-rate-layer. After that, there is a second SimpleRNN layer with 64 hidden units activated by ReLU. The final Dense layer consists of 10 neurons that use the Softmax activation function to generate class probabilities; these neurons stand in for the ten CIFAR-10 classes. The dropout rate in this layer is 0.30. Equation (2) is used to calculate the network's output.

$$y_t = \text{Softmax}(W_{hy}h_t + b_y) \quad (2)$$

where b_y is the output bias vector and W_{hy} is the hidden-to-output weight matrix.This model was trained using the Adam optimizer over the training period of 30 epochs with a learning rate of 0.001 and a batch size of 64. 20% of the training data goes into validation, which helps keep an eye on how well the model is learning. Start with zero bias settings and the Glorot Uniform initializer for the network weights. To make sure the results are consistent every time, use a random seed of 42 to mix up the training sets before each iteration. The objective of optimization is to minimize the categorical cross-entropy loss function, which is obtained from Equation. (3):

$$L = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{3}$$

where C stands for the overall count of classes, y_i is the actual label for class i, and $\hat{y}_i$ represents the likelihood that class i is true. To thoroughly examine the suggested RNN model's picture classification performance, it is trained on the testing dataset and then tested on that dataset using performance metrics.

G. Evaluation Metrics

The proposed model was tested for its performance with various classification metrics. The TP, FP, TN, and FN counts were calculated with the use of a confusion matrix. The F1, REC, ACC, and PRE were all calculated using these variables, giving a thorough evaluation of the model's classification performance:

Accuracy: A metric for evaluating the trained model's performance in making predictions in comparison to the entire dataset (input samples). The Equation for it is (4):

$$Accuracy = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \tag{4}$$

Precision: The accuracy and confidence level (ACC) of a model's predictions is the ratio of the number of positive instances accurately predicted to the total number of positive examples. One such representation is Equation (5):

$$Precision = \frac{\text{TP}}{\text{TP+FP}} \tag{5}$$

Recall: This metric measures the accuracy of positive event predictions as a percentage of the total number of occurrences that ought to have been positive. It may be expressed mathematically as Equation (6):

$$Recall = \frac{\text{TP}}{TP+FN} \tag{6}$$

F1 score: The harmonic mean of PRE and REC is what it is. In other words, it helps to balance PRE and REC. Mathematically, it is given as Equation. (7):

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{7}$$

H. Experiments Setup

All of the tests were executed on a Windows 11 Home 64-bit machine with a 32 GB RAM stick, an Intel Core i7-14700F processor, and an NVIDIA GeForce RTX 5060 graphics processing unit. Used Python 3.10.19 to build and test the proposed model using the PyTorch 2.11.0.dev20260211+cu130 framework and CUDA 13.0 for GPU acceleration.

4. Results and Discussion

The experimental findings and performance analysis of the proposed RNN model for image visual categorisation using the CIFAR-10 dataset are presented in this part. A number of measures are used to evaluate the model. These include learning and validation performance, ACC, PRE, REC, confusion matrix analysis, and loss convergence. The outcomes show that the suggested method is effective, resilient, and can generalize to correct picture classifications across 10 object types.

Table II shows how well the suggested RNN model classified data from the CIFAR-10 dataset. Strong learning power and decent generalization were shown by the model's 98.1% training ACC, 97.1% validation ACC, and 97.2% testing ACC, with just a little performance gap across the datasets. Consistent with this, show consistent convergence and little overfitting in training with losses of 0.10, 0.17, and 0.18 correspondingly. An ACC of 97.0%, REC of 97.6%, and F1 of 97.4% were the high classification metrics attained by the suggested model on the testing set, demonstrating its efficacy in correctly identifying pictures across all ten CIFAR-10 classes. Figure 4 shows the findings, which prove that the suggested RNN model offers dependable and strong visual categorization performance for images.

TABLE II. CLASSIFICATION RESULTS OF PROPOSED MODEL FOR IMAGE VISUAL CLASSIFICATION

Metric	Training	Validation	Testing
Accuracy (%)	98.1	97.1	97.2
Loss	0.10	0.17	0.18
Precision (%)	98.0	97.2	97.0
Recall (%)	98.2	97.4	97.6
F1-Score (%)	98.1	97.3	97.4

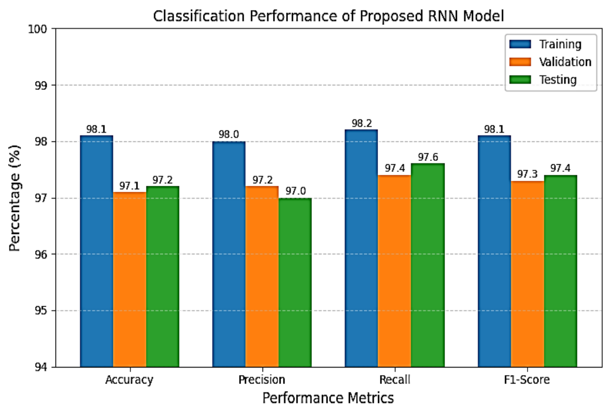

Fig. 4. Bar graph of classification performance of Proposed Model for the visual image classification

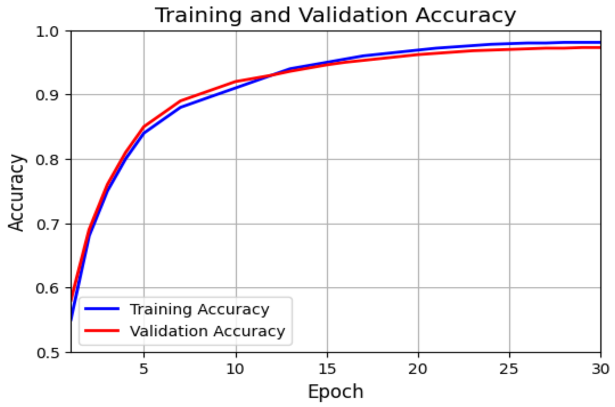

Fig. 5. Accuracy Curve for the Proposed Model

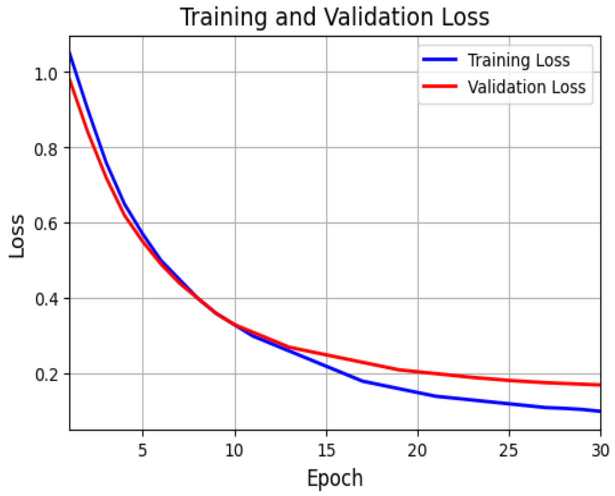

Fig. 6. Loss Curve for the Proposed Model

The suggested RNN model's training and validation performance over 30 epochs are displayed in Figures 5 and 6, respectively. With a small gap of just 0.8%, the ACC curves show high generalization and low overfitting, attaining 98.1% training ACC and 97.3% validation ACC. Training loss falls from 1.05 to 0.10 and validation loss falls from 0.98 to 0.17, indicating ongoing model convergence in a similar vein as the loss curves consistently decline. The findings show that the RNN model is successful; it gets 97.2% of the tests right on the CIFAR-10 dataset.

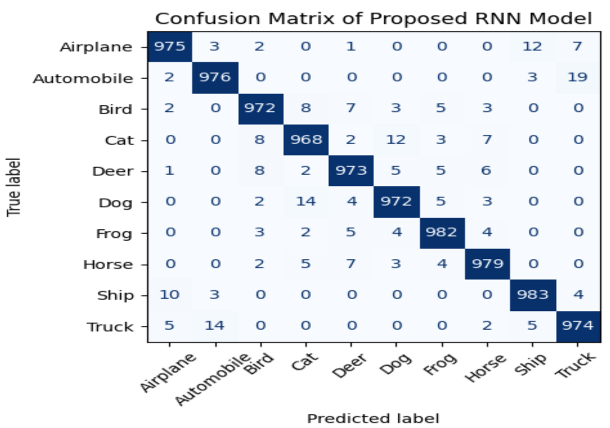

Fig. 7. Confusion matrix of the proposed Recurrent Neural Network

The suggested RNN model's confusion matrix on the CIFAR-10 dataset is shown in Figure 7. The majority of samples are correctly classified, as indicated by the high values along the major diagonal, and there are few misclassifications across visually similar classes like cat-dog, deer-horse, automobile-truck, and aeroplane-ship. The results show that the proposed method is effective and trustworthy when it comes to visual picture classification.

A. Comparative Analysis

Table III presents the results of a comparison of the proposed RNN model's performance with those of other DL models already available for use in order to assess the model's efficacy. Evaluate the proposed model using F1, REC, ACC, and PRE, and compare it to CNN, ResNet50, and MobileNetV3. In comparison to ResNet50 (94.1% ACC) and MobileNetV3 (95.49% ACC), the CNN model only managed 75.7% ACC. But the suggested RNN model outperformed all others with a 97.2% testing ACC, 97.0% REC, and 97.4% F1. In addition, the model demonstrated consistent learning with little overfitting and good generalizability, achieving 98.1% training ACC and 97.1% validation ACC. These outcomes prove that the suggested RNN is superior to the current models and offers a solid alternative for picture visual categorization.

TABLE III Comparison of Different Deep Learning Models for Image Visual Classification

Model	Accuracy	Precision	Recall	F1-score
CNN[24]	75.7	75.9	75.7	75.6
ResNet50 [25]	94.1	95.4	94.6	94
MobilenetV3[26]	95.49	-	-	95.33
Proposed RNN	97.2	97	97.6	97.4

In comparison to deeper and more resource-intensive models, the lightweight design with two SimpleRNN layers offers an efficient compromise between computational complexity and ACC. The implications of this work are significant for various computer vision applications, including automated image recognition, intelligent surveillance, autonomous systems, and visual analytics. Based on the results, RNN-based architectures are a viable option for efficient visual feature learning in image classification tasks. This opens the door to future advancements in hybrid CNN-RNN architectures, attention mechanisms, and advanced recurrent models for large-scale image analysis.

B. Limitations and Future Work

A number of caveats to the suggested method need to be considered. As a small-scale benchmark dataset with low-resolution pictures and limited item categories, CIFAR-10 is the only dataset used to evaluate the model. Therefore, the generalization capability of the proposed RNN model for more complex real-world image classification tasks with larger, diverse, and high-resolution datasets requires further investigation. Additionally, the study focuses on a single RNN-based architecture, and the performance comparison is limited to selected existing DL models. The suggested method validated in future studies by testing its scalability and resilience on bigger and more difficult datasets including CIFAR-100, ImageNet, Caltech-101, and Tiny ImageNet. In order to enhance classification performance and create more versatile visual recognition systems, it is possible to investigate and evaluate more sophisticated models such as CNN, Vision Transformers (ViT), EfficientNet, DenseNet, and hybrid CNN-RNN architectures.

5. Conclusion

Image classification is a key technology in the study and progress of modern technology and science. This research presented an RNN-based framework for image visual classification using the CIFAR-10 dataset. The proposed model effectively utilized sequential feature learning to extract meaningful representations from image data and perform multi-class classification. The model achieved consistent and reliable classification performance through the combination of appropriate preprocessing approaches, data augmentation, and recurrent learning. A testing ACC of 97.2% was produced by the experimental evaluation of the proposed approach, proving that it effectively recognized several item categories within the dataset. The results indicate that RNN-based architectures can be a promising alternative for visual classification tasks. However, the current study is limited by the use of a single dataset and a single model architecture. Exploring advanced architectures such as CNN, Vision Transformer (ViT), EfficientNet, DenseNet, and hybrid CNN-RNN models will enhance scalability, robustness, and real-world applicability in future work. Evaluating the proposed approach on larger and more diverse datasets like CIFAR-100, ImageNet, Caltech-101, and Tiny ImageNet will also be undertaken.

REFERENCES

1. A. Shakya, M. Biswas, and M. Pal, "Parametric study of convolutional neural network based remote sensing image classification," Int. J. Remote Sens., vol. 42, no. 7, pp. 2663–2685, Apr. 2021, doi: 10.1080/01431161.2020.1857877.

2. J. Yang et al., "MedMNIST v2 - A large-scale lightweight benchmark for 2D and 3D biomedical image classification," Sci. Data, vol. 10, no. 1, p. 41, Jan. 2023, doi: 10.1038/s41597-022-01721-8.

3. M. Naseri et al., "A new cryptography algorithm for quantum images," Optik (Stuttg)., vol. 171, pp. 947–959, Oct. 2018, doi: 10.1016/j.ijleo.2018.06.113.

4. J. Maurício, I. Domingues, and J. Bernardino, "Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review," 2023. doi: 10.3390/app13095521.

5. M. N. Paralath, "Bridging Astronomical Imagery and Text with Advanced Multi-Modal AI," 2025 IEEE 7th Int. Conf. Comput. Commun. Autom., pp. 1–8, 2025.

6. R. Rudra, R. Lingam, S. K. Ravva, and S. A, "A Generalized Deep Learning Approach for Cross-Crop Plant Disease Detection Using the Plant Village Dataset," J. Mach. Comput., vol. 5, no. 3, pp. 1592–1605, Jul. 2025, doi: 10.53759/7669/jmc202505126.

7. P. Sharma, Y. P. S. Berwal, and W. Ghai, "Performance analysis of deep learning CNN models for disease detection in plants using image segmentation," Inf. Process. Agric., 2020, doi: 10.1016/j.inpa.2019.11.001.

8. M. Naseri, M. Abdolmaleky, F. Parandin, N. Fatahi, A. Farouk, and R. Nazari, "A New Quantum Gray-Scale Image Encoding Scheme*," Commun. Theor. Phys., vol. 69, no. 2, p. 215, Feb. 2018, doi: 10.1088/0253-6102/69/2/215.

9. S. Mukherjee, "AI-Driven Deep Learning System for Alzheimer's Disease (AD) Detection Via MRI Imaging," in 2026 International Conference on Intelligent Computing, Networks, and Security (IC-ICNS), Bhubaneswar, India: IEEE, 2026, pp. 1–6, March. doi: 10.1109/IC-ICNS68863.2026.11537765.

10. M. Mittal, "Exploring Generative Adversarial Networks (GANs) For Realistic Image Synthesis," Int. J. Innov. Res. Comput. Commun. Eng., vol. 06, no. 02, Feb. 2018, doi: 10.15680/IJIRCCE.2018.0602133.

11. B. Jeganathan, "Exploring the Power of Generative Adversarial Networks (GANs) for Image Generation: A Case Study on the MNIST Dataset," Int. J. Adv. Eng. Manag., vol. 7, no. 1, pp. 21–46, Jan. 2025, doi: 10.35629/5252-07012146.

12. R. Nair, S. Vishwakarma, M. Soni, T. Patel, and S. Joshi, "Detection of COVID-19 cases through X-ray images using hybrid deep neural network," World J. Eng., vol. 19, no. 1, pp. 33–39, Feb. 2022, doi: 10.1108/WJE-10-2020-0529.

13. S. Singamsetty, "AI-powered Satellite Image Processing for Global Air Traffic Surveillance Techniques Using NCNN–EGSA Optimization Techniques," in Machine Learning Based Air Traffic Surveillance System Using Image Processing,

Emerald Publishing Limited, 2026, pp. 179–197. doi: 10.1108/978-1-80592-062-520251010.

14. S. A. D. Poovaiah, "Benchmarking provable resilience in convolutional neural networks: A study with Beta-CROWN and ERAN," Int. J. Sci. Res. Arch., vol. 7, no. 2, pp. 834–845, 2022.

15. S. Heidari, M. M. Abutalib, M. Alkhambashi, A. Farouk, and M. Naseri, "A new general model for quantum image histogram (QIH)," Quantum Inf. Process., vol. 18, no. 6, p. 175, Jun. 2019, doi: 10.1007/s11128-019-2295-5.

16. S. Baamonde, M. Cabana, N. Sillero, M. G. Penedo, H. Naveira, and J. Novo, "Fully automatic multi-temporal land cover classification using Sentinel-2 image data," Procedia Comput. Sci., vol. 159, pp. 650–657, 2019, doi: 10.1016/j.procs.2019.09.220.

17. M. N. Paralath, "Towards Interpretable and Attentive Deep Visual Perception Through Iterative Refinement in Convolutional Neural Networks," in 2025 International Conference on Advances in Next-Gen Computer Science (ICANCS), IEEE, Nov. 2025, pp. 1–6. doi: 10.1109/ICANCS65819.2025.11376668.

18. R. S. S, P. A, K. B, and S. S. Mohamed, "EfficientNet-B0 Based Multi-Class Kidney Abnormality Classification from CT Images Using Transfer Learning," in 2026 Twelfth International Conference on Bio Signals, Images, and Instrumentation (ICBSII), IEEE, Mar. 2026, pp. 1–6. doi: 10.1109/ICBSII69710.2026.11479194.

19. V. K, V. S, and K. Gayathri, "Real-Time Defect Classification in Manufacturing using Machine Learning and Image Processing," in 2026 International Conference on Visual Analytics and Data Visualization (ICVADV), IEEE, Feb. 2026, pp. 1707–1709. doi: 10.1109/ICVADV67766.2026.11470348.

20. F. Saadati-Monem, A. Mansouri, A. Mahmoudi-Aznaveh, and H. Motamednia, "The Effect of Training Strategy on Image Classification Using Representations of Learning-Based Image Compression," in 2025 7th International Conference on Pattern Recognition and Image Analysis (IPRIA), IEEE, Sep. 2025, pp. 1–6. doi: 10.1109/IPRIA68579.2025.11263702.

21. H. Yuan, Z. Xiong, and B. Chen, "Phenotypic Analysis and Variety Classification of Waxy Corn Seeds Based on Machine Vision," in 2024 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML), IEEE, Nov. 2024, pp. 912–915. doi: 10.1109/ICICML63543.2024.10958153.

22. C. V. Krishna Prasad, V. Amudha, and S. S. H. R, "Accuracy improvement of SAR image classification method using Greedy Hierarchical Learning in comparison with Support Vector Machine," in 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), IEEE, May 2023, pp. 1–7. doi: 10.1109/ACCAI58221.2023.10200978.

23. J. Yang, J. Xue, X. Feng, C. Song, and Y. Hao, "Tire Pattern Image Classification using Variational Auto-Encoder with Contrastive Learning," in 2022 IEEE International Conference on Visual Communications and Image Processing (VCIP), IEEE, Dec. 2022, pp. 1–5. doi: 10.1109/VCIP56404.2022.10008835.

24. P. Kiran Mehta, R. Jain, and M. Batra, "Image Classification using CNN and CIFAR-10," Int. Res. J. Eng. Technol., pp. 317–326, 2024.

25. M. H. Mohiuddin and L. Tamilselvan, "IDedupNet: A MobileNetV3-Based Deep Learning Framework for Efficient Image Deduplication in Cloud Computing Environments," Informatica, vol. 49, no. 13, Feb. 2025, doi: 10.31449/inf.v49i13.7514.

26. T. Shahriar, "Comparative Analysis of Lightweight Deep Learning Models for Memory-Constrained Devices," pp. 1–22, 2025.