IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 6, Issue 1, September 2026 Impact Factor: 8.2

A Hybrid Deep Learning Architecture for Image Classification Across Diverse Visual Recognition Applications

Dr Chintal Kumar Patel
Associate Professor, CSE
Geetanjali Institute of Technical Studies, Udaipur, Rajasthan
chintal.patel@gits.ac.in

Abstract: *The core computer vision problem of picture categorization has several domain-specific applications. This paper's goal is to talk about the significance of picture categorization in modern technology and society, as well as its ideas, techniques, and applications. Many computer vision systems rely on image classification, where images are automatically assigned to a predetermined category, based on the information they provide in the image. This study trains a hybrid deep learning system to efficiently recognize photos using the Caltech-256 dataset. The system makes use of both Bidirectional Long Short-Term Memory (BiLSTM) and Gated Recurrent Units (GRU). Dataset preprocessing includes standardization, label encoding, damaged image identification, and duplicate reduction. After separating the dataset into a train and test set, the following step is to extract features from each. The proposed GRU+BiLSTM model is evaluated in comparison to AlexNet, MobileNetV2, EfficientNet, and CNN models using ACC, PRE, REC, and F1-score (F1). Results for ACC (98.9%), PRE (99.1%), REC (99.3%), and F1 (100%) were better using the suggested method compared to the top deep learning models, according to the available experimental data. Findings validate the proposed hybrid design's provision of a strong and efficient.*

Keywords: Computer Vision Image Classification, Deep Learning (DL), Image Processing, Object Classification, Object Detection.

I. INTRODUCTION

Computer vision relies on image classification, which involves labeling input pictures according to their visual features [1]. Many practical uses rely on it, such as autonomous vehicles, robots, industrial inspection, medical image analysis, surveillance systems, remote sensing, and content-based image retrieval [2][3]. The rapid developments of digital technologies and imaging devices have produced huge amounts of visual information every day, which has led to the need for accurate and automatic image classification methods. The ability of intelligent computers to comprehend visual information in many application domains is made possible by image classification, which is why it is considered a top issue in computer vision among image processing and pattern analysis [4].

The use of SIFT, HOG, texture descriptors, and color histograms for feature extraction in the past has been greatly superseded by more modern classification techniques [5]. These methods were successful on relatively simple data sets, but required extensive handcrafted features and domain knowledge [6][7]. They struggle to generalize when presented with changes in the appearance of objects, lighting conditions, scale, the perspective from which the object is viewed, and background complexity [8]. However, these traditional methods were limited when dealing with larger and more complex image datasets in terms of their ability to capture representative visual features and achieve high classification ACC [9].

The advent of AI and digital image processing has been a game-changer for computer vision. In recent years, image classification is an important technique in many applications, including healthcare, agriculture, intelligent transportation, environmental monitoring, manufacturing and security [10]. Commonly used image classification methods involve image preprocessing steps including resizing, normalization and augmentation, image feature extraction, and predicting image categories [11]. Improving classification ACC and processing efficiency, as well as developing more robust ways to capture rich visual information, has been motivated by the growing complexity of visual data [12][13].

The use of ML and DL approaches has led to considerable progress in image classification in recent years. Automated learning of picture characteristics from extensive datasets is enabled by these methods, which subsequently unleash advanced pattern recognition capabilities. The need for human feature engineers is eliminated by DL, which automatically generates hierarchical feature representations from raw images. In contrast, ML techniques acquire discriminative features by utilizing previously acquired features [14][15][16]. They have proven to be highly successful in various visual recognition tasks. Based on the above, the study introduces A Hybrid DL Architecture for Image Classification Across Diverse Visual Recognition Applications, where Hybrid DL methods are used to enhance the feature representation and classification performance for robust and generalized image recognition [17][18].

A. Motivation and Contribution

The purpose of this research is to fill need for accurate and reliable image classification in various visual recognition applications. The existing image classification technologies have been found to have low performance in complex visual patterns and under different of object classes. A hybrid GRU-BiLSTM DL architecture aims at improving feature learning and classification ACC. The planned solution should be a powerful, efficient, and broadly applicable solution for contemporary computer vision applications. This research makes several important contributions, as follows:

- Proposed hybrid GRU–BiLSTM DL model for precise image classification in visual recognition applications.
- Test the proposed picture classification algorithm on the dataset of 30,607 photographs and 257 object types from Caltech.
- Cleaned data, encoded labels and normalized data for improving the quality of the images.
- Conducted a comprehensive ACC, PRE and REC analysis and F1 analysis to evaluate the proposed model.
- Proposed an effective DL framework for image classification for various visual recognition applications.

B. Justification and Novelty

The increasing demand for accurate image classification across diverse visual recognition applications requires models that can effectively learn complex visual patterns while maintaining high classification performance. Even the standard DL models have shortcomings in representing full feature dependency or even in ensuring uniform ACC for various object types. They suggest that a combination of these properties of GRU and BiLSTM may be necessary to overcome these concerns. The proposed method, combined with the implementation of feature extraction and image preprocessing, improves the accuracy of the classification and achieves good performance in image recognition applications, showing its novel and effective characteristics.

C. Structure of the Paper

The paper is organized as follows: Section II presents related work. The materials and methods along with data preparation, model, and architecture are described in Section III. A quantitative evaluation and an explainability analysis of the results are incorporated into Section IV, which includes the experimental setup, results, and discussion. The paper concludes with Section V, which offers prospective directions.

II. LITERATURE REVIEW

A comprehensive literature study and analysis of current research on visual recognition picture categorisation is carried out to lay a solid groundwork for the suggested method's development.

He and Chen (2026) proposes a self-supervised visual model optimization framework based on contrastive learning, which generates multi-view samples through data augmentation strategies, extracts features using an encoder network, and employs a dual projection head architecture to achieve collaborative optimization of similarity loss and representation learning. The results indicate that the proposed method outperforms the SimCLR method by 2.213 percentage points and supervised learning methods by 3.826 percentage points in image classification tasks on the CIFAR-10 datasets. The rate of the Top-1 ACC is 92.347%. [19].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, September 2026
Impact Factor: 8.2

Saadati-Monem et al. (2025) Computer vision tasks such as object identification, classification, and segmentation are not facilitated by methods such as JPEG, H.264, or HEVC, which are optimized for human viewing but not for deep learning models. Not only does this approach enhance ACC for picture classification by stabilizing the training process and utilizing the extracted features, but it also preserves and enhances the compression network's performance in comparison to conventional methods. The proposed methodology has been demonstrated to enhance performance by an average of 7.99% [20].

Mahadevan et al. (2024) Investigates into use of ML techniques to image recognition tasks that pertain to security. The preprocessing methods consist of OpenCV, Pillow, and TensorFlow, while the features are extracted using Wavelet, CNN, and LBP. There are other options for picture recognition algorithms, but SVM achieves the best ACC at 90%. Jobs like employe verification and suspect identification in public areas rely on robots that can interpret visual data, which is the aim of this research [21].

Krishna Prasad, Amudha and R (2023) offered to compare and contrast supported vector machines with a new greedy hierarchical learning approach for SAR image categorization. The Kaggle learning platform is used to extract a total of 20 dataset samples. Separate sets, designated as training and testing, make up the dataset. Of the twenty datasets available, ten are designated for testing and ten for training purposes. The G power is used to determine the sample size after correcting the alpha (0.05) and power (80%). The results show that in the MATLAB simulation, SVM obtained a classification ACC of 80.41% and new greedy hierarchical learning reached 93.08%. The significance level that was produced by the SPSS analysis was 0.009 (p 0.05) [22].

Punnium, Rujikietgumjorn and Rattanatamrong (2022) built TH-UniformDB, a database containing 10,000 photos of a single individual donning a Thai uniform across nine distinct professions, 1,000 images for each of these disciplines. Experimental results show that by combining grayscale and RGB image characteristics, the recognition ACC of the classic DNN model may be improved by 3.15 to 10.15% [23].

Nehvi, Dar and Assad (2021) The suggested approach outperformed the stock model, which achieved an inaccurate 70%. The training section of the dataset only comprised 320 photos for all categories, with 80 images per category. This research contributes to the expanding body of evidence that transfer-based strategies can be effectively employed to address computer vision issues, even in the presence of limited data [24].

Table I summarizes recent work on image classification for visual recognition, including the models proposed, the datasets used, the significant findings, and the challenges that arose in each study.

Table I. Recent Research on Image Classification in Visual Recognition Applications

Authors	Implementation	Datasets	Findings	Limitations & Recommendation
He and Chen (2026)	Proposed a self-supervised visual model based on contrastive learning with dual projection heads and joint similarity loss for label-efficient image representation learning.	CIFAR-10 (60,000 images, 10 classes)	Achieved 92.35% Top-1 accuracy, improving 3.83% over supervised learning and 2.21% over SimCLR.	Evaluated only on CIFAR-10. Future work includes testing on larger and more complex datasets and reducing computational cost.
Saadati-Monem et al. (2025)	Developed an interactive learning-based image compression framework integrated with an image classification network for compressed-domain feature learning.	Standard image classification datasets (not explicitly specified in the paper summary)	Improved image classification performance by an average of 7.99% while maintaining compression efficiency.	Limited evaluation on broader vision tasks. Future work includes extending the framework to object detection and semantic segmentation.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, September 2026 Impact Factor: 8.2

Mahadevan et al. (2024)	Compared multiple machine learning classifiers using OpenCV, TensorFlow, Wavelet, CNN, and LBP feature extraction for security image recognition.	Security image dataset	SVM achieved highest classification accuracy of 90%.	Focused mainly on conventional machine learning methods. It suggests integrating DL architectures for improved robustness.
Krishna Prasad, Amudha and R (2023)	Proposed a Greedy Hierarchical Learning approach and compared it with SVM for SAR image classification.	Kaggle SAR image dataset (20 samples)	Greedy Hierarchical Learning achieved 93.08% accuracy compared with 80.41% for SVM.	Very small dataset limits generalization. Future work should validate the method on larger SAR datasets and real-world scenarios.
Punnium, Rujikietgumjorn and Rattanatamrong (2022)	Proposed occupation recognition by combining RGB and grayscale image features within deep neural networks.	TH-UniformDB (10,000 images, 10 occupation classes)	Improved recognition accuracy by 3.15%–10.15% over traditional deep neural networks.	Limited to occupation recognition. Future work includes evaluating on diverse image datasets and multimodal feature fusion.
Nehvi, Dar and Assad (2021)	Used transfer learning using a pre-trained Inception-v3 for visual object recognition with limited training data.	Custom dataset (320 training images, 4 object classes)	Achieved approximately 90% accuracy, outperforming a conventional CNN (70%).	Dataset is very small and contains limited object categories. Future work includes testing on larger datasets and modern transfer learning architectures.

Research gaps: Although visual recognition performance has been greatly enhanced by current image classification approaches, many of these methods are either tailored to particular datasets or depend on a single DL architecture, which severely restricts their capacity to accurately capture varied object categories and intricate visual patterns. Other studies address feature optimization, image compression, or transfer learning, but not to improve overall classification ACC over large-scale and multi-class datasets. In order to enhance feature representation, classification accuracy, and generalization performance, a strong hybrid DL architecture is indeed necessary. This study presents a hybrid GRU-BiLSTM model that reliably classifies images in different visual recognition applications and aims to address this issue.

III. MATERIALS AND METHOD

The proposed methodology uses the Caltech-256 dataset, where images are preprocessed through duplicate removal, corrupted image detection, label encoding, and StandardScaler normalization to improve data quality. Using a hybrid GRU-BiLSTM DL model, hierarchical visual characteristics are retrieved and classified from the preprocessed photographs after the dataset is divided into a training set of 70% and a testing set of 30%. To validate the model's effectiveness in image classification, use a confusion matrix along with REC, ACC, PRE, and F1 to evaluate its performance. The suggested flowchart for visual recognition image classification is shown in Fig. 1.The next section describes each step in the proposed methodology in detail:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, September 2026
Impact Factor: 8.2

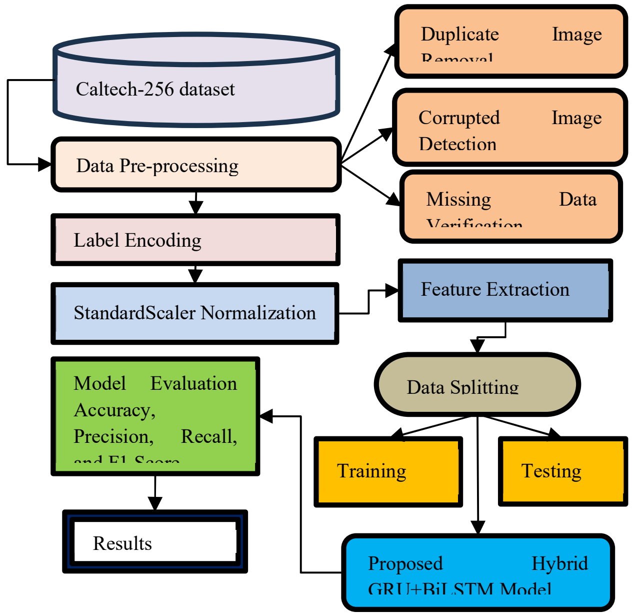

Fig.1. Proposed flowchart for Image Classification for Visual Recognition

A. Data Collection

The Caltech-256 dataset consists of a collection of 30,607 images gathered from Google Image Search and PicSearch.com, and classified into 257 object categories. All images were thoroughly checked by human annotators for quality and class ACC, and the visualizations of the dataset for class distribution and other characteristics are shown below:

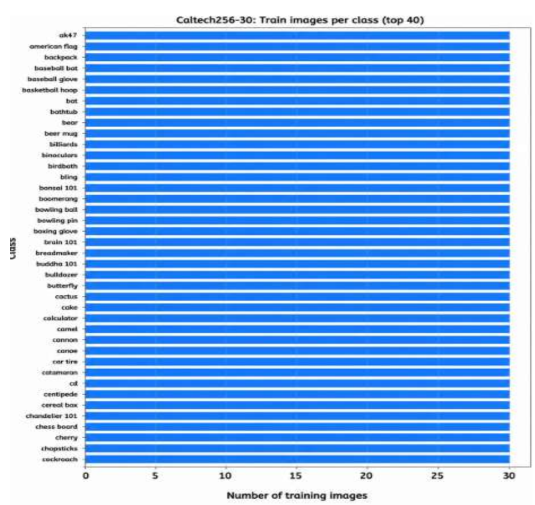

Fig 2. Image Class Distribution

The distribution of the training images for the top 40 classes in the Caltech-256 dataset is shown in Fig 2. The number of photos per class is 30, and there is no class bias. This makes reliable categorization of the pictures possible.

The diversity of the visual classes for training and testing is illustrated in Fig 3 with representative sample images from each of object categories in Caltech-256 dataset. Images are taken under various lighting conditions, background information, and perspectives for a comprehensive evaluation of image classification models.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, September 2026
Impact Factor: 8.2

Fig 3. Sample images

B. Data Pre-processing

Data concatenation, Data cleansing and feature engineering, Using Caltech 256 dataset. The pre-processing stages were missing value imputation, duplicate removal, noise elimination and redundant value removal, data Lebling and normalization. A quick rundown of the most critical preprocessing procedures is provided here:

- **Duplicate Image Removal:** Redundant images and model bias during training are eliminated with duplicate image removal.
- **Corrupted Image Detection:** Eliminating corrupted or unreadable image files to ensure the use of valid images for model training.
- **Missing Data Verification:** Removal of images that lack adequate data or metadata to maintain consistency and quality of the dataset.
- **Label Encoding:** Label encoding transforms categorical data, like attack types or class labels, into numerical values suitable for ML algorithms.

C. StandardScaler Normalization

The data is scaled to be uniform with the help of the StandardScaler() function. This makes the resulting distribution mean = 0 and standard deviation = 1. The formula for this transformation is (1), which shows that one must divide the average of all observations by their standard deviation:

$$z = \frac{x-\mu}{\sigma} \quad (1)$$

z stands for the changed value of the feature, x for the initial value of each descriptor, μ for the mean of the feature, and σ for the standard deviation in the dataset.

D. Feature Extraction

Automatic extraction of hierarchical visual elements from preprocessed pictures, including edges, textures, forms, and object patterns, is achieved using convolutional layers. Additional hierarchical visual properties that may be extracted from input photos using convolutional layers include forms, textures, contours, edges, and object structures. Convert the feature maps that were extracted into feature vectors that are ordered sequentially, all while keeping the crucial spatial information. The sequences are inputted into the hybrid DL network.

E. Data Splitting

A random split of 70% for train and 30% for test was applied to picture dataset. The proportionality of this divide enables fair and accurate model training as well as impartial evaluation of classification performance on unseen images.

F. Proposed Hybrid GRU+BiLSTM Model

A DL-based model for visual recognition picture categorization using a BiLSTM and a GRU is presented in this paper. The suggested hybrid model draws on the strengths of GRU and BiLSTM networks to capture the interdependencies in sequential data, whether they are short-term or long-term. GRUs are also capable of mitigating the vanishing gradient problem and are computational efficient, whereas BiLSTM can process a sequence from both the forward and backwards direction, which is useful for gaining a better context understanding and improving prediction ACC. This is a useful

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, September 2026 Impact Factor: 8.2

combination when modelling time series or sequence problems where the ability to capture the temporal structure is important. These are the changes to the GRU:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (2)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \quad (3)$$

$$\bar{h}_t = tanh(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h) \quad (4)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \bar{h}_t \quad (5)$$

This suggested hybrid GRU+BiLSTM model excels at sequential data workloads thanks to its balanced computational efficiency and predictive effectiveness. The GRU component speeds up the learning process while maintaining the essential memory, and the BiLSTM component is used to capture the bidirectional relationships for more reasonable ACC in forecasting. The experimental findings show that compared to individual GRU and LSTM models, this architecture can reduce errors and increase convergence speed. During the 100-epoch training phase, the hybrid model is trained with a loss function that combines the Adam optimizer with categorical cross-entropy. There are 32 participants in the group, and the learning rate is set at 0.001. Avoid overfitting by using a dropout of 0.5 and achieve consistent convergence and high classification ACC using 128 hidden layer units.

G. Evaluation Metrics

A battery of industry-standard categorization measures is used to evaluate the suggested model. A confusion matrix displays total no. of correctly and incorrectly classified instances should be generated initially using equations (6–9). The results of the confusion matrix were used to calculate an evaluation metric known as True Negatives (TN), False Positives (FP), True Positives (TP), and True Negatives (FN). A value that is utilized to compute ACC, PRE, REC, and F1 is classification ability of proposed model:

$$Accuracy(ACC) = \frac{\text{TP+TN}}{\text{TP+Fp+TN+FN}} \quad (6)$$

$$Precision(PRE) = \frac{\text{TP}}{\text{TP+FP}} \quad (7)$$

$$Recall(REC) = \frac{\text{TP}}{TP} \quad (8)$$

$$F1 - score(F1) = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

The overall classification ACC can be calculated, for example, by dividing the total number of input samples by the sum of all predictions. To calculate ACC of positive predictions, use PRE, which is ratio of TP forecasts to entire number of predictions. The REC, or percentage of properly predicted cases, indicates how sensitive a model is to positive examples. F1, a harmonic mean of PRE and REC, is measured between 0 and 1, with closer to 1 indicating better categorization ability. Beneficial for evaluating unbalanced datasets fairly, since it helps to ensure that each.

IV. RESULTS AND DISCUSSION

The system's trials employ a 16 GB of RAM, an Intel CoreTM i3-10100F CPU operating at 3.60 GHz, and an NVIDIA graphics card. Training and evaluation conducted on Windows 11 64-bit OS, ensuring stability and efficiency in ML model development.

A. Experimental Results

The proposed model was on Caltech-256 Dataset using key performance indicators listed in Table II. The proposed hybrid GRU+BiLSTM model has been able to achieve 98.9% ACC, which shows a very good overall classification performance. It also achieved a 99.1% PRE in positive predictions, a measure of ACC. In addition, model achieved 99.3% REC and 98.6% F1, indicating that good performance in accurately identifying and recalling various image categories.

Table II. Performance of the proposed model for image classification

Matrix	TEST	TRAIN
ACC	0.989	100
PRE	0.991	100
REC	0.993	100

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, September 2026
Impact Factor: 8.2

F1-score	0.986	100

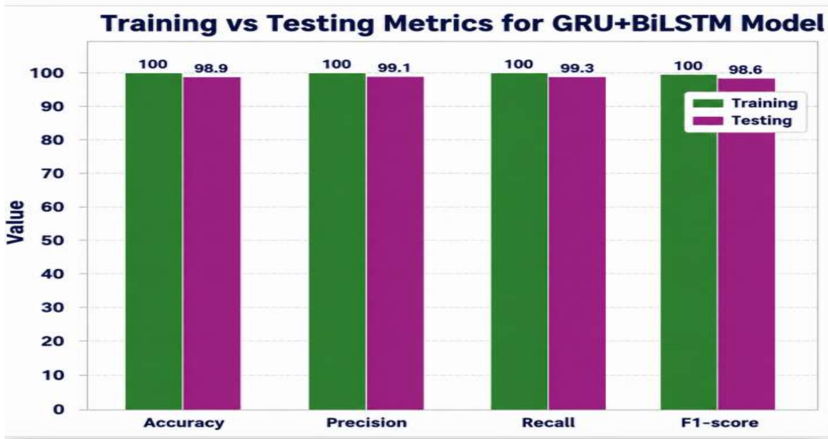

Fig 4. Train and Test performance comparison of proposed Model

In Fig 4, a bar chart, the efficacy of the hybrid GRU+BiLSTM model that was planned was compared during training and testing using ACC, PRE, REC, and F1. In training, the model achieved 100% ACC; in testing, it achieved 98.9% ACC, 99.1% PRE, 99.3% REC, and 98.6% ACC, demonstrating excellent classification performance with limited fluctuation between the training and testing sets.

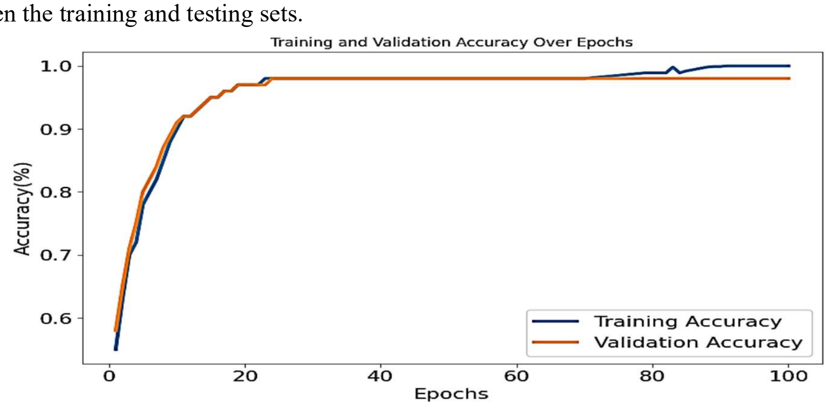

Fig 5. Accuracy curve for proposed Model

The proposed model for train and val over 100 epochs is presented at ACC in Fig 5. Indicative of quick convergence and efficient learning, the curves blend quickly at the outset and stabilize at values around 98-99% ACC. Close proximity of the training and test curves is an advantageous training effect and does not indicate overfitting.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, September 2026
Impact Factor: 8.2

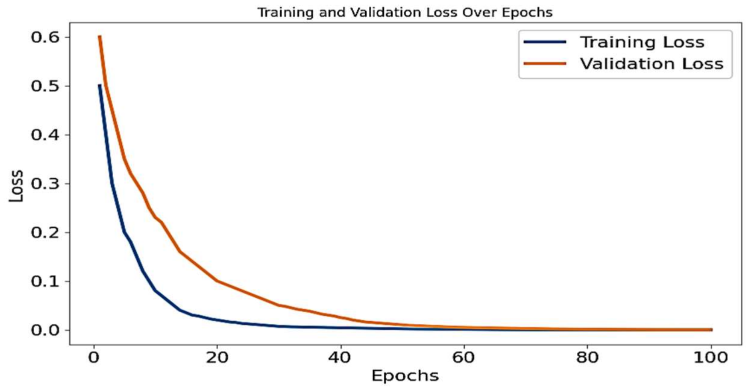

Fig 6. Loss Curve for Proposed Model

Train and Test losses following 100 iterations of proposed model are shown in Fig 6. The loss curves are steep and converge to 0.0 in the first few epochs, which suggests stable optimization and learning. There are minimal overfitting and excellent generalization when the training loss and validation loss values are so close to one another.

B. Comparative Analysis

Table III lists existing DL models that are compared to the proposed model to ensure its efficacy. This evaluation takes a look at the suggested hybrid GRU+BiLSTM model, as well as AlexNet, MobileNetV2, EfficientNet, And CNN. The proposed model outperformed all of others in terms of maximum ACC (98.9%), PRE (99.1%), REC (99.3%), and F1 (98.6%). Both discrimination and classification performance are improved by the hybrid design, according to the results. Furthermore, it is reliable and may be used for many other kinds of visual identification tasks. When compared to other DL-based models, the GRU+BiLSTM model performs better overall.

Table III. Comparison of Different Deep Learning Models for Image Classification in Visual Recognition

Model	Accuracy	Precision	Recall	F1-score
AlexNet[25]	40.13	-	-	-
MobileNetV2[26]	90.37	74	72	73
EfficientNet [27]	84.25	90.62	80.53	82.44
CNN[28]	90.31	94.52	94.94	94.64
Proposed hybrid GRU+BiLSTM model	98.9	99.1	99.3	98.6

C. Performance Discussion

The proposed hybrid GRU+BiLSTM model demonstrated its effectiveness in learning complex visual features for image recognition tasks with the outstanding classification ACC. The enhanced performance is ascribed to the fact that GRU and BiLSTM networks are capable of representing both the immediate and distant relationships between features. In a number of visual recognition applications, outcomes validate that suggested model is an effective and appropriate option for picture categorization.

V. CONCLUSION AND FUTURE STUDY

The goal of visual recommendation systems is to provide users with tailored suggestions for media (photos, movies, items, etc.) depending on their tastes, behaviors, and other pertinent characteristics. This approach is applicable in a wide variety of settings, such as online video streaming, social media, online shopping, and other visual recognition systems. In this research, examine several DL models for picture categorization using a comprehensive comparison analysis. Out of all the models tested, the CNN model, AlexNet, EfficientNet, and MobileNetV2 all had worse classification ACC than the suggested hybrid GRU+BiLSTM model's 98.9%. Acquiring both short-term and long-term feature dependencies enables suggested approach to get accurate and reliable visual recognition, as demonstrated by its better performance. These results shows that suggested hybrid architecture is powerful and effectual technique for picture categorization in variety of visual identification applications.This study's drawback is that the suggested hybrid GRU–BiLSTM model was only tested on the Caltech-256 dataset, and its comparatively high computing resource requirements may prevent it from being applied to other large-scale picture datasets. The next steps involve the validation of the model with a wider range of benchmark datasets, enhancing computational efficiency with the development of advanced architectures, and expanding the framework to real-time image classification and other computer vision tasks.

REFERENCES

1. S. Devulapalli, A. Potti, R. Krishnan, and M. S. Khan, "Experimental evaluation of unsupervised image retrieval application using hybrid feature extraction by integrating deep learning and handcrafted techniques," Mater. Today Proc., vol. 81, pp. 983–988, 2023, doi: 10.1016/j.matpr.2021.04.326.
2. I. Ahmed, M. Ahmad, F. A. Khan, and M. Asif, "Comparison of Deep-Learning-Based Segmentation Models: Using Top View Person Images," IEEE Access, vol. 8, pp. 136361–136373, 2020, doi: 10.1109/ACCESS.2020.3011406.
3. G. Cheng, X. Xie, J. Han, L. Guo, and G.-S. Xia, "Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 13, pp. 3735–3756, 2020, doi: 10.1109/JSTARS.2020.3005403.
4. G. Li, Y. Yang, X. Qu, D. Cao, and K. Li, "A deep learning based image enhancement approach for autonomous driving at night," Knowledge-Based Syst., vol. 213, p. 106617, Feb. 2021, doi: 10.1016/j.knosys.2020.106617.
5. H. K. Yadav, "Design and Implement Machine Learning Based Predictive Analytics for Road Safety and Accident Prevention," in 2026 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), IEEE, Mar. 2026, pp. 1–6. doi: 10.1109/IATMSI68868.2026.11465965.
6. X. Gao, M. Zhang, and J. Luo, "Low-Light Image Enhancement via Retinex-Style Decomposition of Denoised Deep Image Prior," Sensors, vol. 22, no. 15, p. 5593, Jul. 2022, doi: 10.3390/s22155593.
7. S. Gite, A. Mishra, and K. Kotecha, "Enhanced lung image segmentation using deep learning," Neural Comput. Appl., vol. 35, no. 31, pp. 22839–22853, Nov. 2023, doi: 10.1007/s00521-021-06719-8.
8. M. Mittal, "Understanding Natural Language Processing (NLP) Techniques: From Text Analysis to Language Generation," Int. J. Res. Comput. Appl. Inf. Technol., vol. 7, no. 2, pp. 2784–2792, Dec. 2024, doi: 10.34218/IJRCAIT_07_02_213.
9. D. Jin, H. Zheng, Q. Zhao, C. Wang, M. Zhang, and H. Yuan, "Generation of Vertebra Micro-CT-like Image from MDCT: A Deep-Learning-Based Image Enhancement Approach," Tomography, vol. 7, no. 4, pp. 767–782, Nov. 2021, doi: 10.3390/tomography7040064.
10. K. Venkata and R. K. Koppireddy, "Revolutionizing Identity Verification: AI-Driven Digital Identity Solutions for a Secure and Seamless Future," Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol., vol. 11, no. 1, pp. 2814–2824, Feb. 2025, doi: 10.32628/CSEIT251112301.
11. M. N. Paralath, "Bridging Astronomical Imagery and Text with Advanced Multi-Modal AI," in 2025 IEEE 7th International Conference on Computing, Communication and Automation (ICCCA), IEEE, Nov. 2025, pp. 1–8. doi: 10.1109/ICCCA66364.2025.11325649.
12. J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Jun. 2009, pp. 248–255. doi: 10.1109/CVPR.2009.5206848.
13. C. Djellali, M. Adda, and M. T. Moutacalli, "A Data-Driven Deep Learning Model to Pattern Recognition for Medical Diagnosis, by using Model Aggregation and Model Selection," Procedia Comput. Sci., vol. 177, pp. 387–395, 2020, doi: 10.1016/j.procs.2020.10.052.

14. Y. K. Dwivedi et al., "Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," Int. J. Inf. Manage., vol. 57, p. 101994, Apr. 2021, doi: 10.1016/j.ijinfomgt.2019.08.002.
15. P. Kaur, G. Singh, and P. Kaur, "Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification," Informatics Med. Unlocked, vol. 16, p. 100151, 2019, doi: 10.1016/j.imu.2019.01.001.
16. S. Tyagi and A. Tyagi, "Evaluating the Performance with Deep Learning Techniques in Cybersecurity for Effective Threat Detection," 2026, pp. 437–452. doi: 10.1007/978-981-96-8687-2_31.
17. S. D. Pawar, A. Tiwari, A. K. Singh, A. Banerjee, S. Lokhande, and R. A. Bagate, "Remote Health Monitoring System using Secure Communication Techniques," in 2023 3rd International Conference on Intelligent Technologies, CONIT 2023, 2023. doi: 10.1109/CONIT59222.2023.10205905.
18. N. Manakitsa, G. S. Maraslidis, L. Moysis, and G. F. Fragulis, "A Review of Machine Learning and Deep Learning for Object Detection, Semantic Segmentation, and Human Action Recognition in Machine and Robotic Vision," Technologies, vol. 12, no. 2, p. 15, Jan. 2024, doi: 10.3390/technologies12020015.
19. Z. He and Z. Chen, "Optimization of Deep Visual Models for Image Recognition Using Self-Supervised Learning," in 2026 9th International Conference on Advanced Algorithms and Control Engineering (ICAACE), IEEE, Mar. 2026, pp. 1835–1838. doi: 10.1109/ICAACE69793.2026.11508859.
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. N. Mahadevan, T. Swetha, N. S. Kumudini, R. Yogendra, K. S. N. Sree, and T. Sajitha, "Enhancing Image Recognition for Security Applications: A Machine Learning Approach," in 2024 Second International Conference on Advances in Information Technology (ICAIT), IEEE, Jul. 2024, pp. 1–6. doi: 10.1109/ICAIT61638.2024.10690431.
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. V. Punnium, S. Rujikietgumjorn, and P. Rattanatamrong, "Classifying Thai Occupation from Images using Deep Learning with Grayscale Feature Extractor," in 2022 19th International Joint Conference on Computer Science and Software Engineering (JCSSE), IEEE, Jun. 2022, pp. 1–6. doi: 10.1109/JCSSE54890.2022.9836300.
24. A. Nehvi, R. Dar, and A. Assad, "Visual Recognition of Local Kashmiri Objects with Limited Image Data using Transfer Learning," in 2021 International Conference on Emerging Techniques in Computational Intelligence (ICETCI), IEEE, Aug. 2021, pp. 49–52. doi: 10.1109/ICETCI51973.2021.9574047.
25. J. Luo and D. Hu, "An Image Classification Method Based on Adaptive Attention Mechanism and Feature Extraction Network," Comput. Intell. Neurosci., vol. 2023, no. 1, Jan. 2023, doi: 10.1155/2023/4305594.
26. S. Albawi, L. Kamil Almajmaie, and A. J. Abboud, "A Comparative Study of Deep Learning Methods-Based Object/Image Categorization," Iraqi J. Electr. Electron. Eng., vol. 21, no. 1, pp. 168–177, Jun. 2025, doi: 10.37917/ijeee.21.1.17.
27. Livinus Ifunanya Umeaduma, "Survey of image classification models for transfer learning," World J. Adv. Res. Rev., vol. 21, no. 1, pp. 373–383, Jan. 2024, doi: 10.30574/wjarr.2024.21.1.0006.
28. M. A. Rafidison et al., "Image Classification Based on Light Convolutional Neural Network Using Pulse Couple Neural Network," Comput. Intell. Neurosci., vol. 2023, no. 1, Jan. 2023, doi: 10.1155/2023/7371907.