Journal of Information Systems Engineering and Management
2025, 10(25s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

EnsembleDRM Model for Multiclass Image Classification in Deep Learning

Rana J. AL-Sukeinee[1], Noor M. Al-Moosawi[2], Jenan A. Alhijaj[2], Raidah S. Khudeyer[2], Zaid Ameen Abduljabbar[3,4], Vincent Omollo Nyangaresi [5, 6]

[1]Department of Physic, College of Science, University of Basrah, Basrah, 61004, Iraq.

[2]Department of Computer Science, College of Computer Science and Information Technology, University of Basrah, Basrah, 61004, Iraq.

[3]Department of Computer Science, College of Education for Pure Sciences, University of Basrah, Basrah, 61004, Iraq.

[4] Department of Business Management, Al-Imam University College, Balad 34011, Iraq

[5]Department of Computer Science and Software Engineering, Jaramogi Oginga Odinga University of Science and Technology, Bondo 40601, Kenya.

[6]Department of Applied Electronics, Saveetha School of Engineering, SIMATS, Chennai, Tamil Nadu, 602105, India.

* Corresponding author's Email: rana.jabbar@uobasrah.edu.iq

ARTICLE INFO

Received: 29 Dec 2024

Revised: 14 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

Convolutional neural networks (CNNs) have performed exceptionally well in various computer vision tasks. Previously, researchers relied on feature extraction and then classification. With CNNs, feature extraction and classification are performed in a single step. Therefore, incorporating a set of convolutional neural networks, known as an ensemble, can help increase the effectiveness of their behavior. This paper presents an architecture that is the result of collaborative efforts, consisting of three types of CNN: Densenet121, Resnet101, and MobileNetV2. We also present experimental results using the Cifar10 and Cifar100 datasets, achieving impressive classification accuracy of 99% for Cifar10 and 86% for Cifar100. This case study contributes to deep learning optimization and benefits researchers and practitioners looking for optimal approaches in various computer vision applications. We also evaluate the scalability and robustness of our proposed method in the context of different CNN structures by using more than one model.

Keywords: deep learning, convolutional neural networks, densenet121, resnet101, and mobilenetv2.

INTRODUCTION

In depth learning, the computer vision and image recognition industries have been entirely transformed by convolutional neural networks, a powerful type of artificial neural networks [1]. These networks have achieved impressive accuracy in a number of fields during the past ten years and are especially created to analyze visual data effectively [2]. Researchers are widely interested in image processing, feature extraction, and classification because of their influential role in serving society. Convolutional neural networks, in particular, have demonstrated a high level of efficiency in image processing, and as a result, researchers are constantly developing new models employing deep learning. CNN's advantage in handling images was demonstrated by their ability to learn from the depth and the absence of feature extraction beforehand [3]. Several methods have been suggested to improve CNN's performance, one of which is ensemble learning. To achieve better performance and generalizability than training convolutional networks, a single Ensemble learning method trains multiple base learners and combines their expectations [4]. Ensemble learning is one effective technique in the field of artificial intelligence for resolving a variety of issues [5]. Combining several models to act as experts or classifiers is the outcome. The performance of models in a variety of tasks, including classification and prediction, is enhanced by this kind of learning. Additionally, it effectively lowers the likelihood that poorly performing models would be incorrectly selected and increases confidence in the model's decision. In ensemble learning, mixing and matching learners yields various benefits regarding statistics, computation, and representation [6].

To support ensemble learning in making the best decisions and achieving high accuracy, a combination of multiple pre-trained models that have achieved state-of-the-art image classification in this field with high capacity and high accuracy in extracting features have been developed. Today there are a lot of real-life applications where multi-class classification problems are used such as: image classification. In multi-class classification problem appear imbalanced in datasets an imbalanced dataset means the number of the majority class is much more than the minority class. The Ensemble learning technique is the one solution for this problem, where with ensemble learning combines more than one classifier to improve the performance of the models and obtains higher classification accuracy than a single classifier.

RELATED WORK

In 2020[7], FELIPE O. GIUSTE et al. presented a study combining several image feature sources from both deep learning and manual methods. This work aimed to improve the classification of pictures from the Cifar10 dataset. Four methods were used to generate image features, which utilize deep learning. After FCNN architecture was trained using image feature sets, the final accuracy evaluations were obtained using testing data. The results were 53% and 59% classification accuracy for HOG (histogram of oriented gradients) and pixel intensities, 60% and 93.43% accuracy (a Cifar10 (ICIFAR-VGG) and VGG16 with ImageNet trained weights for enhanced picture classification.

Ross Wightman et al. [8] focused on experimenting with different training techniques and their impact on the performance of the model. ResNet-50 is the proposed network that was used, a vanilla ResNet-50 achieves 80.4% top-1 accuracy at 224×224 resolution on ImageNet-val. He concluded a close relationship between high results and fine-tuning of parameters.

In 2021 [9], Mingxing Tan et al., this paper suggests neural networks (EfficientNetV2), a smaller, faster convolutional neural network for image classification. This architecture suggested a more effective progressive learning technique than scaling and regularizing the image to further speed up training EfficientNetV2 is up to 11x faster than the previous EfficientNet and up to 6.8x smaller (achieving 87.3% top-1 accuracy on mageNet ILSVRC2012, uitilizing the same computer resources code). The spread of the SARS-CoV-2 virus (COVID-19) has damaged global healthcare systems, resulting in the death of almost a million people globally. Thus, it has set the global economy back to a standstill. In 2021[10], Rohit Kundu and coworkers demonstrated a system that combines ensemble deep learning (VGG-11, GoogLeNet, SqueezeNet v1.1, Wide ResNet-50-2) with a Sugeno fuzzy. This system utilized chest CT-scan images to identify patients with COVID-19 and non-COVID groups, and the results outstripped the latest technologies using the same SARS-COV-2 dataset.

Kaiqi Zhao et al. [11] suggested a methodology for classifying images that could lessen the issue of resource scarcity and lengthy training times. Using pre-trained CNN and cloud architecture networks, ResNet and VGG networks were used to create a lightweight model with less depth and fewer parameters. The Cifar 10 dataset was used for testing the model. In 2022[12], Shuai Zhao el al. Introduced a model in which one network is separated into multiple smaller ones concerning its parameters and regularization elements. This can give better results (accuracy and efficiency) than merely widening the width of the network. These small networks are training together to increase diversity and see different views using the same data. In addition, these networks can learn with one another while undergoing a co-training process. As a result, these small networks can accomplish better performance than one huge network with few parameters. In several applications, like natural image recognition and language modeling, it has been demonstrated that increasing the model, compute budget, and data scale in the pre-training forcefully improves model generalization and transfer.

In 2022 [13] Mehdi Cherti el al., presented system combines X-ray chest imaging datasets and its open access to achieve a scale for medical imaging equivalent to ImageNet-1k, frequently used for pre-training in the natural image field. Due to the more extensive pre-training range and significant improvements for intra-field natural and medical transfer within the field in the full shot system, the advancement is unclear for the smaller targets and the Low Shot regime. The summary of the study indicates that a significantly expanded model and general, medical, and typical natural image source data scale during pre-training can allow high feature out-of-domain transfer to field medically-specific targets, reducing reliance on medically-specific source data that are frequently unavailable in an experience. Another research [14] aims to adjust the values of the filters used within the convolutional layers by finding the importance of the input to the output. Networks will be used for the training process. Ensemble techniques and transfer learning were used by Rasool Fakhir et al. [15] to create a model for medical diagnosis. The networks AlexNet, ResNet-50, and VGG-16 were used as feature extraction. The suggested approach produced state-of-the-art outcomes when compared to peers, demonstrating the effectiveness of the employed ensemble with pre-trend model. In addition, transfer learning was used to reinstate the ideal network weights for Inception, ResNet v2 (TL-Inception) and VGG16 (TL-VGG) which significantly improved classification performance (by 85% and 90.74%, respectively).

Our effort entails building a transfer learning architecture with several subnetworks, such as Densenet121, Resnet101, and MobileNetV2, In addition to the suggested CNN. The architecture plays a major role in the functions of feature extraction, image recognition, and multi-categorization (up to 100 different types). In this study, ensemble learning is also employed to enhance the performance of the suggested model by amalgamating predictions from several models, hence augmenting the precision and adaptability of the predictions.

Table 1. Summarize More Relative Research.

	Research	Year	Dataset	Method	Accuracy
7	CIFAR-10IMAGE CLASSIFICATION USING FEATURE ENSEMBLES	2020	Cifar10	CIFAR-VGG TL-Inception	60% ,93.43% , 85 90.74%
8	ResNet strikes back: An improved training procedure in time	2021	Cifar10 Cifar100	vanilla ResNet-50	97% 80.4%
11	Enabling Deep Learning on Edge Devices through Filter Pruning and Knowledge Transfer	2022	Cifar10	WRN-28-10, ResNet-34, VGG-16	90%
14	BETA-RANK: A ROBUST CONVOLUTIONAL FILTER PRUNING METHOD FOR IMBALANCED MEDICAL IMAGE ANALYSIS	2023	Cifar10 cifar100	VGG-16 ResNet56	93.13%, 72.75%, 92.09, 67.11
15	Ensemble Deep Learning Technique for Detecting MRI Brain Tumor	2024	MRI images	Ensemble learning AlexNet, ResNet-50, VGG-16	99.16%

MOTIVATION AND KEY CONTRIBUTION

Finding and developing an architecture that can effectively contribute to accurately distinguishing between different classes through images is the main goal of the study. The main contributions can be summarized in the following points:

- Propose architecture based on deep CNNs that can effectively extract low-level image features. The CNN architecture adopted in the study is the most recently used in literature.
- Ensemble learning is employed to avoid singularity and complexity in training findings on the efficacy of a single classifier alone. This improves decision accuracy and reduces errors.
- Two datasets, Cifar10 with ten classes and Cifar100 with 100 classes, were tested to verify the efficiency of the architecture and the techniques.
- Hyperparameter adjustment and appropriate training methods, such as the Adam optimizer with the learning scheduler technique used to optimize the learning rate value for faster convergence and to achieve high accuracy with a more generalized model.

TRANSFER LEARNING

Traditional CNN networks begin by extracting features from the convolutional portion, typically comprising many convolutional layers, pooling layers, and activation functions. Subsequently, the fully linked section receives the extracted features for the ultimate classification procedure [16].

Large datasets directly impact the effectiveness of CNN training, potentially requiring longer training cycles. Transfer learning can assist in overcoming this by using the knowledge acquired from another discipline. In transfer learning, open-source CNN models that have already been trained and tuned by developers can be used with optimized weights as a starting point to train the network much faster and achieve better performance [17].

We examine the necessary background for the three models used in our proposed approach: DenseNet121, ResNet101, and MobileNetV2.

DenseNet121

DenseNet121 is the state-of-the-art architecture of CNN, which was initially suggested by Huang et al. The core idea behind Densenet-121 is to reuse features in order to reduce the number of trainable parameters and improve computational efficiency, allowing for a deeper network [18]. The network is organized into dense blocks composed of several layers to reduce the vanishing gradient issue. Each layer in a dense block is connected to its predecessors, and the input of the next layer is merged with the previous layer's output. Thus, DenseNet-121's primary benefit is that it handles vanishing gradients, which has several advantageous effects: it decreases the training load of deep learning models, uses fewer parameters than other common deep learning models, and allows for the reuse of attributes [19]. Figure 1 displays architecture of a DenseNet121.

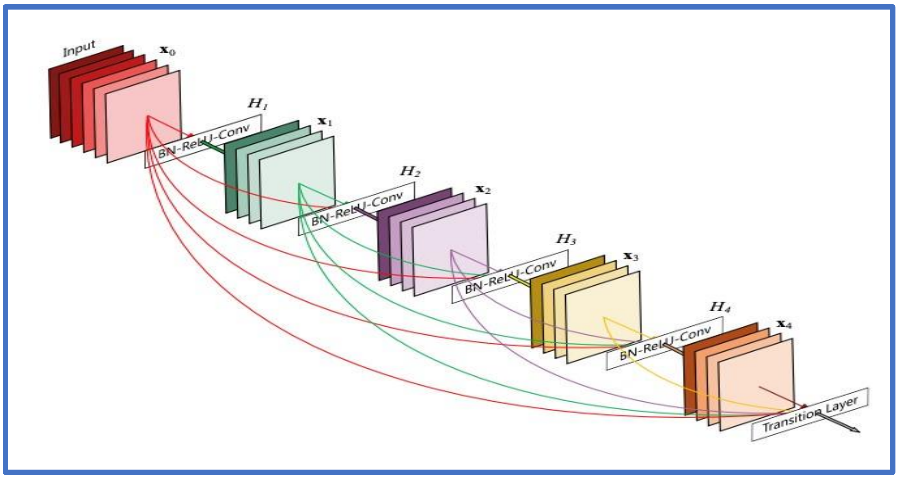

Figure 1. DenseNet121architecture.

ResNet101

ResNet101 depicts the residual network and effectively contributes to computer vision tasks like image classification and object detection [20]. In addition, ResNet101 represents the deepest suggested design for ImageNet. It contains 104 convolutional layers and 33 blocks of layers, and 29 of these blocks are used directly in earlier layers. Due to the gradient that can quickly dwindle to zero, depth Networks experience vanishing gradient troubles [21]. Gradients in ResNets101 can stream over skip connections and propagate back to the previous layers, as shown in Table **2** figure 2 displays a residual building block that can be determined as:

$$y = F(x, \{W_i\}) + x \qquad (1)$$

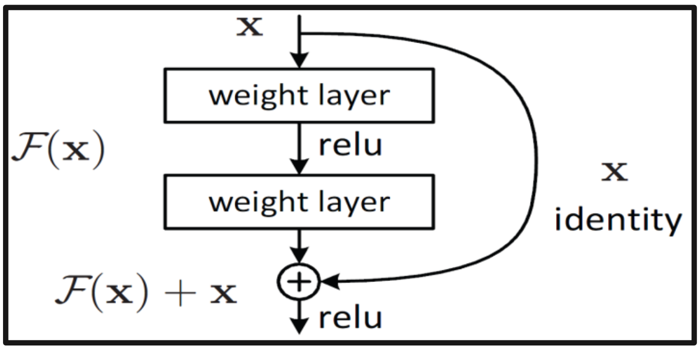

Figure 2. shows residual Block [22].

Table 2. Architecture of ResNet-101.

Layer Name	Output Size	101-Layer
conv1	112×112	7×7,64, stride 2
conv2	56×56	3×3 max pool, stride 2 $\begin{bmatrix} 1\times 1,64 \\ 3\times 3,64 \\ 1\times 1,256 \end{bmatrix} \times 3$
conv3	28×28	$\begin{bmatrix} 1\times 1,128 \\ 3\times 3,128 \\ 1\times 1,512 \end{bmatrix} \times 4$
conv4	14×14	$\begin{bmatrix} 1\times 1,256 \\ 3\times 3,256 \\ 1\times 1,1024 \end{bmatrix} \times 23$
conv5	7×7	$\begin{bmatrix} 1\times 1,512 \\ 3\times 3,512 \\ 1\times 1,2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-D fc, softmax
FLOPs		7.6×10^9

MobileNetV2

One of the models used in the proposed system is MobileNet V2. It is a deep-learning algorithm explicitly designed for mobile applications and image processing [23]. The MobileNetV2 model involves 32 filtered convolutional layers and 19 residual bottleneck layers. It is built on an inverted residual structure where residual connections connect the bottleneck layers. Lightweight depthwise convolutions are utilized in the middle extension layer as a source of nonlinearity to purify features. In addition, MobileNetV2 can decrease Convolutional layers' temporal complexity and spatial complexity of convolutional layers; for this reason, MobileNetV2 is a suitable choice for applications that need real-time processing yet have limited computing power [24]. Figure 3 depicts MobleNetV2.

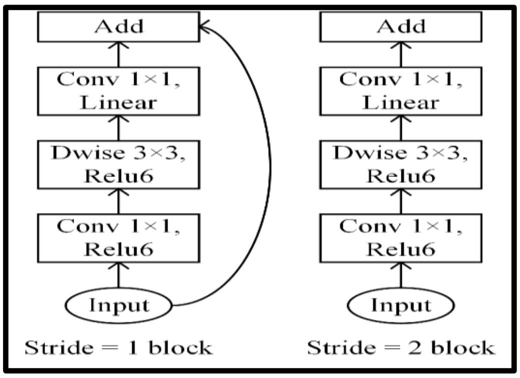

Figure 3. MobileNetV2 architecture [25].

DATASET DISTRIBUTION

To assess the effectiveness of our approach, we selected the Cifar10 and Cifar100 datasets for their widespread use in the deep learning community, manageable dataset sizes, and diverse content. This enabled us to effectively assess the generalizability of the current work across different domains.

- Cifar10

One of the common datasets for machine learning is Cifar10, which is used for many applications. It is well-documented and translucent. This dataset contains 60000 photos, 32x32 pixels in size, divided into ten classes. In addition, the Cifar10 dataset is divided into three sets: 40,000 images for the training phase, 10,000 for the validation phase, and 10,000 images for the testing phase [26].

- Cifar100

The other database utilized in this study is Cifar100. It shares the same features as Cifar10, but it stands out with its 100 classes, a significant difference from the 10 classes in Cifar10. The Cifar100 dataset is also divided into three sets: 40,000 images for training, 10,000 for validation, and 10,000 images for testing [27].

METHODOLOGY

Data Preprocessing

When working with images, preprocessing is a crucial step to accentuating features and enhances model performance. The following steps represent the processing mechanism used:

- Image Resizing: since DenseNet121, ResNet101, and MobileNetV2 pre-trained models, images were resized into 128 widths x 128 heights to fit input layers, reduce resource waste, and speed up processing.

- Image normalization: Min-Max Normalization provides a uniform scale in a range of [0…1], which speeds up the convergence process in the gradient descent algorithm and thus shortens the training time. The formula for calculating Min-Max scaling is as follows:

- $$X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad (2)$$

where X_{max}-represent the maximum pixel value in image , X_{min} –minimum pixel values in image.

EnsembleDRM Architecture

To obtain the suggested ensembleDRM model with an appropriate level of accuracy several steps are performed as follows:

Step1: Three pre-trained CNN models—DenseNet121, ResNet101, and MobileNetV2—that were described in the preceding section were integrated as the initial step in creating the ensembleDRM model. Only the feature extraction portion of each chosen CNN model was utilized as a subnetwork; the fully linked portion was cut off.

Step2: The subnetworks from step1 were combined into an extensive neural network with multiple heads, to increase the submodels' likelihood of learning features from every input during the training phase.

Step3: As an extra branch of the model, a basic 3-Conv2D layers network was built to increase the network's efficacy. Three Convolution layers with sixty-four filters were used to recover the regional and local feature map. A pooling layer (max pooling) was used after convolution layers to return only important features and reduce the feature maps size.

Step4: The results of each model are combined using a simple neural network designed using two fully connected (Dense) layers.

Step5: In the fully connected part, a hidden layer was defined using ReLU as the activation function. The ReLU activation function is used to help mitigate the vanishing gradient problem during model training and enable CNN models to learn more complex relationships in data. The ReLU formula is explained as follows, if the input is positive, the ReLU activation function outputs it directly; otherwise, it will output zero.

$$f(x) = \max(0, x) \qquad (3)$$

Step6: The output layer consisting of 10 or 100 nodes to create a single vector of probabilities that the model predicts based on training on Cifar10 or 100. Softmax was used to distinguish between the classes and generate probabilistic predictions. The EnsembleDRM model is explained in Table 3. Fig 4 shows a graphical illustration of the proposed model.

Table 3. Detalled explanation of EnsembleDRM

Branch 1	**Branch2**	**Branch3**	**Branch4**
DenseNet121	ResNet101	MobileNetV2	Conv2D(64,)
			BN()
			MaxPooling2D
			Dropout(0.25)
			Conv2D(64,)
			BN()
			MaxPooling2D
			Dropout(0.25)
			Conv2D(64,)
			BN()
			MaxPooling2D
			Dropout(0.25)
GlobalAveragePooling2D()	GlobalAveragePooling2D()	GlobalAveragePooling2D()	Flatten()
Dense(256,)	Dense(256,)	Dense(256,)	Dense(256,)
BN()	BN()	BN()	BN()
Dropout(0.5)	Dropout(0.5)	Dropout(0.5)	Dropout(0.5)
Dense(10,)	Dense(10,)	Dense(10,)	Dense(10,)
Ensemble vote			

EnsembleDRM Classifier

The ensemble technique is applied to determine the classes of images. Each image's classifiers are integrated, and the best voting mechanism determines the categorization. Every classifier in ensemble learning predicts a label for the class using the majority voting process, which we have employed. The target label for this image is given to the class that received the most votes or the highest prediction. The following formula determines the outcome of the majority vote:

$$\tilde{y} = argmax\left(br_1(y_t^1), br_2(y_t^2), \ldots, br_n(y_t^n)\right) \qquad (4)$$

Where $br\ \ (y_t\ )$ indicates the category that got majority of votes and the base learner DenseNet121, ResNet101, indicated by (1, 2, ..., n).

Implementation and Training Technique

The proposed model was first trained and evaluated on Cifar10 with ten classes, later modified to classify Cifar100 with 100 categories. The Kaggle platform with a GPU P100 graphics card was used to train and test the model.

Training and validation data are passed to the model during the training phase. Validation data is used to check the model's progress and contribute to fine-tuning the parameters after each training cycle. Since the networks are pre-trained, ideal weights were used as a starting point instead of training from scratch. This stage is crucial for speeding up training because the base layers can recognize simple structures in images, including lines and curves, and have already learned some prior knowledge.

The error was calculated using categorical cross-entropy, and the Adam optimizer with the learning scheduler technique was used to optimize the learning rate value for faster convergence.

Thirty training cycles were used to train the models, and callback approaches include:

- ModelCheckPoint: to store the optimal weights for the model that produced the least amount of loss on the validation data.
- EarlyStop: to end training early if performance did not improve after five training cycles to save time and prevent overfits.
- ReduceLROnPlateau: if the validation loss does not improve, the learning rate is decreased.

Table 4. EnsembleDRM Model Training Parameters.

Input shape	(128, 128, 3)
Batch size	64
epochs	30
compile	Adam

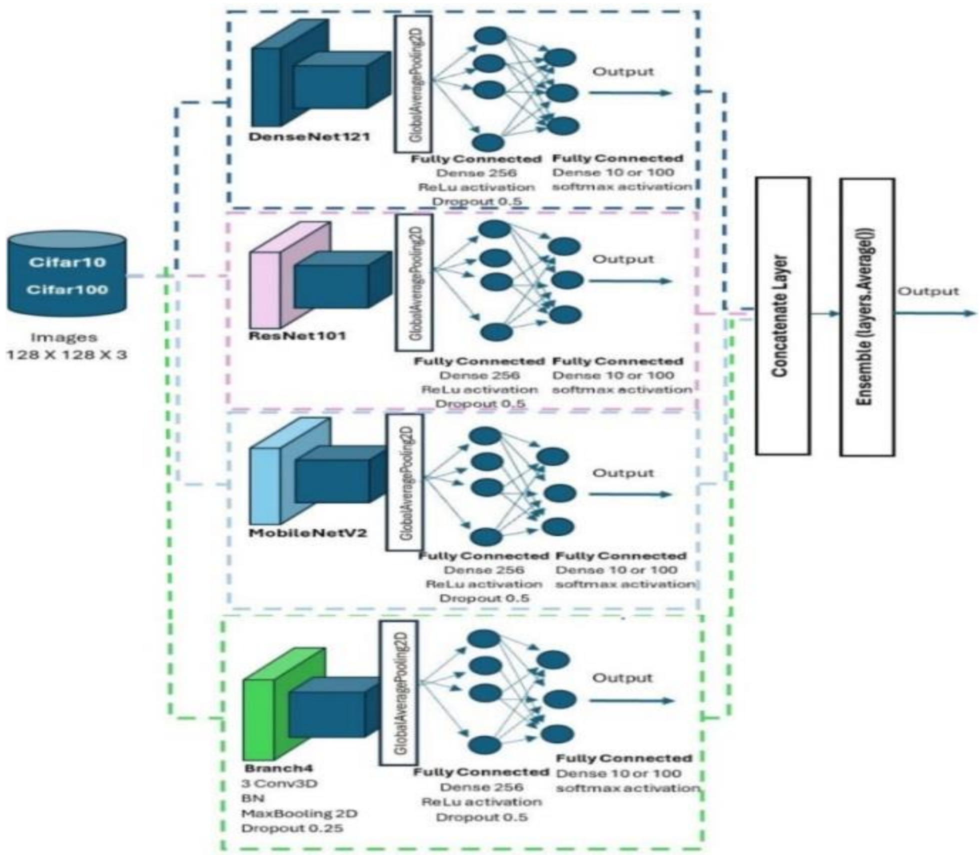

Figure 4. The proposed EnsembleDRM Model Architecture.

RESULTS AND ANALYSIS

Each selected model was trained individually to obtain their respective outcomes in the first training step. After completing the training process and reaching an appropriate accuracy on the training data, the test data was used to verify the accuracy of the results and the efficiency of the models. Performance measures including accuracy, recall, precision, and F1 score are used. However, the following formulas mathematically quantify performance measures for a given number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN):

$$Accuracy = \frac{Total\ correct\ classification}{Total\ classification} = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

Sensitivity=TP/(TP+FN) (7)

$$F1\ score = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity} \quad (8)$$

of the individual networks, which is a critical stage in the training process. The results revealed that the final model outperformed the individual models, achieving impressive accuracy of up to 99%.

Tabel 5. Different Pre-Training Model Results Compared with the EnsembleDRM Model.

Method	Cifar10 Accuracy	Sensitivity	Precision	F1 score
DenseNet121	93.65%	92.5%	89.8%	91.1%
ResNet101	96%	87.1%	94.4%	90.6%
MobileNetV2	90.9%	88.23%	90.9%	89.54%
EnsembleDRM	99	99	98.68	98.83

We present the results of our approach and compare it to other researchers' methods in terms of their accuracies on the Cifar10 and Cifar100. Table 6 shows the experimental results of our approach compared to other methods on two datasets. Most previous research employed pre-trained networks, which performed exceptionally well in discriminating between Cifar10 and Cifar100 datasets' categories. WRN-28-10, ResNet-34, VGG-16 [8], VGG-16, ResNet56 [7], vanilla ResNet-50 [10], and Convolution deep learning [9] are among the models that were assessed. Our method, the EnsembleDRM model, which combines three pre-trend models with an ensemble learning strategy, has the most fantastic accuracy and beats most other approaches. The EnsembleDRM model achieved 99% accuracy on Cifar10 and 86% accuracy on Cifar100.

These results suggest that the use the ensemble model is highly competitive and may offer superior accuracy compared to other techniques when applied to the Cifar10 and Cifar100. It showcases the effectiveness of combination of many models in enhancing CNNs for image classification tasks.

Table 6. Comparison of Accuracy of the Current Work and Different Models on the Cifar10 Dataset.

Ref	**Method**	**Cifar10 Accuracy**	**Cifar100 Accuracy**
[7]	VGG-16 ResNet v2	93.43%, 90.74%	
[8]	WRN-28-10, ResNet-34, VGG-16	90%	
[9]	Convolution deep learning	73.5%	
[10]	vanilla ResNet-50	97%	85%
[11]	VGG-16, ResNet56	93.96, 93.26,	74.24 73,120
Our method	EnsembleDRM	99	86

CONCLUSION

This paper presents our approach to optimizing CNNs using the many different pre-trained models: Densenet121, ResNet101, and MobileNetV2. As initial attempts, the three types of CNNs were employed to classify multi-class tasks from publicly available Kaggle datasets (Cifar10 and Cifar100), with different CNN structure, we achieving classification accuracy 99% for Cifar10and 86% for Cifar100. The first attempts suffered from overfitting, despite several attempts to reduce it by adding dropout and batch normalization techniques and augmenting data. We think the main reasons are unbalanced data and the weakness of each network's ability to extract sufficient features for the learning process independently from the proposed data. Seamlessly, integrating three types of CNNs helps it in optimization and machine learning. The majority voting method is used in the ensemble learning approach to determine the final output, which aggregates the predictions of several classifiers. Performance proves to be extremely effective, higher with ensemble learning than with individual or conventional learning approaches. Our plan to improve the ability to classify multiclass tasks in the optimization framework involves two main goals. First, we use three different types of CNN architecture and two datasets. Second, we choose the best hyperparameter to achieve more accuracy and make the model easy to adopt in other tasks.

REFRENCES

[1] Abduljaleel, Q. I, Ali and I.H., "Deep Learning and Fusion Mechanism-based Multimodal Fake News Detection Methodologies: A Review," Engineering, Technology & Applied Science Research, vol. 14, pp. 15665-15675, 2024.

[2] R. J. AL-Sukeinee and R. S. Khudeyer, "A Review of Deep Learning and Fuzzy Logic Applications," Artificial Intelligence & Robotics Development, vol. 4, no. 3, pp. 292-313, 2024.

[3] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci and M. & Parmar, "A review of convolutional neural networks in computer," Artificial Intelligence Review, vol. 57, no. 4, p. 99, 2024.

[4] X. Dong, Z. Yu, W. Cao, Y. Shi and Q. & Ma, "A survey on ensemble learning," Frontiers of Computer Science, vol. 14, pp. 241-258., 2020.

[5] I. D. Mienye and Y. & Sun, "A survey of ensemble learning: Concepts, algorithms, applications, and prospects," IEEE Access, vol. 10, pp. 99129-99149, 2022.

[6] L.-H. Li and R. Tanone, "Ensemble Learning based on CNN and Transformer Models for Leaf Diseases Classification," In 2024 18th International Conference on Ubiquitous Information Management and Communication (IMCOM), pp. 1-6, 2024.

[7] F. O. Giuste and J. & Vizcarra, "Cifar-10 image classification using feature ensembles," arXiv preprint arXiv:2002.03846., 2020.

[8] R. Wightman, H. Touvron and H. & Jégou, "ResNet strikes back: An improved training procedure in timm," arXiv preprint arXiv:2110. 00476, 2021

[9] M. Tan and Q. & Le, "Efficientnetv2: Smaller models and faster training. arXiv 2021. arXiv preprint arXiv:2104.00298.," arXiv 2021. arXiv preprint arXiv:2104.00298., 2021.

[10] R. Kundu, P. K. Singh, S. Mirjalili and R. Sarkar, "COVID-19 detection from lung CT-Scans using a fuzzy integral-based CNN ensemble," Computers in Biology and Medicine., vol. 138, p. 104895, 2021.

[11] K. Zhao, Y. Chen and M. & Zhao, "Enabling deep learning on edge devices through filter pruning and knowledge transfer," arXiv preprint arXiv:2201.10947, 2022.

[12] S. Zhao, L. Zhou, W. Wang, D. Cai, T. Lam and Y. & Xu, "Toward better accuracy-efficiency trade-offs: Divide and co-training.," IEEE Transactions on Image Processing, vol. 31, pp. 5869-5880, 2022.

[13] M. Cherti and J. & Jitsev, "Effect of pre-training scale on intra-and inter-domain, full and few-shot transfer learning for natural and X-Ray chest images.," nternational Joint Conference on Neural Networks (IJCNN),IEEE, pp. 1-9, 2022.

[14] M. Homayounfar, M. Koohi-Moghadam, R. Rawassizadeh and V. & Vardhanabhuti, "Beta-Rank: A Robust Convolutional Filter Pruning Method For Imbalanced Medical Image Analysis," arXiv preprint arXiv:2304.07461, 2023.

[15] R. F. Jader, S. W. Kareem and H. Q. Awla, "Ensemble Deep Learning Technique for Detecting MRI," Applied Computational Intelligence and Soft Computing., vol. 2024, no. 1, p. 6615468, 2024.

[16] S. Anwar, S. R. Soomro, S. K. Baloch , A. A. Patoli and A. R. Kolachi, "Performance Analysis of Deep Transfer Learning Models for the Automated Detection of Cotton Plant Diseases.," Engineering, Technology & Applied Science Research., vol. 13, no. 5, pp. 11561-11567, 2023.

[17] Y. Ma, S. Chen, S. Ermon and D. & Lobell, ". Transfer learning in environmental remote sensing," Rmote Sensing of Environment, vol. 103, p. 113924, 2024.

[18] T. S. Arulananth, S. Prakash, R. Ayyasamy, V. Kavitha, P. Kuppusamy and P. & Chinnasamy, "Classification of Paediatric Pneumonia Using Modified DenseNet-121 Deep-Learning Model," IEEE Access, 2022.

[19] S. A. Albelwi, "Deep architecture based on DenseNet-121 model for weather image recognition," International Journal of Advanced Computer Science and Applications, vol. 13, no. 10, 2022.

[20] Q. Zhang, "A novel ResNet101 model based on dense dilated convolution for image classification," SN Applied Sciences, vol. 4, pp. 1-13, 2022.

[21] G. W. Wicaksono and Andreawan, "ResNet101 Model Performance Enhancement in Classifying RiceDiseases," Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 7, no. 2, pp. 345-352, 2023.

[22] Aksoy, B., & Salman, S. O. K. M. "Prediction of Covid-19 disease with resnet-101 deep learning architecture us ing computerized tomography images," Türk Doğa ve Fen Dergisi, vol. 11, no. 2, pp. 36-42, 2022.

[23] K. Dong, Zhou, C., Ruan, Y. and & Li, Y., "MobileNetV2 model for image classification," In 2020 2nd International Conference on Information Technology and Computer Application (ITCA) IEEE, pp. 476-480, 2022.

[24] Indraswari.R., Rokhana, R. and & Herulambang, W., "Melanoma image classification based on MobileNetV2 network," Procedia computer science 197, pp. 198-207, 2022.

[25] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," In Proceedings of the IEEE conference on computer vision and pattern recognition., pp. 4510-4520, 2018.

[26] B. Recht, Roelofs, R., Schmidt, L. and & Shankar, V., "Do cifar-10 classifiers generalize to cifar-10." URL https://arxiv. org/abs., 2018.

[27] R. Zaheer and H. Shaziya, "A Study of the Optimization Algorithms in Deep," In 2019 third international conference on inventive systems and control (ICISC) IEEE., pp. 536-539, 2019.