Journal of Innovative Technology, Vol. 7, No. 2, pp. 1-8, September 2025
http://doi.org/10.29424/JIT.202509_7(2).0001

Using Ensemble Learning to Improve Handwritten Character Recognition Performance

Kuei-Chung Chang [1*], Sheng-Quan Chen [2]

ABSTRACT

This paper proposes a direct and effective approach to improving handwritten character recognition by employing an ensemble learning strategy. By integrating three lightweight convolutional neural network architectures—ResNet, AlexNet, and VGG—through a weighted averaging ensemble, we demonstrate a significant enhancement in prediction accuracy and stability. Even with these relatively simple model structures, the ensemble method effectively harnesses their complementary strengths, leading to improved recognition performance without extensive model optimization. Visualizations from Grad-CAM further illustrate that each base model focuses on distinct features of the input, and through ensemble learning, the combined model mitigates biases and reduces dependency on irrelevant areas such as the background. Our results, validated on the EMNIST dataset, show that the ensemble approach offers a straightforward yet powerful means to enhance both accuracy and robustness in handwritten character recognition tasks, making it particularly suitable for real-world applications where computational efficiency and reliability are also a priority.

Keywords: Handwritten Character Recognition, Ensemble Method

1. INTRODUCTION

Handwritten character recognition is a technology that converts image content into computer text. It can transform handwritten character images into digital formats, such as in applications like mail address recognition, bank check processing, and electronic form filling. Even today, handwritten documents remain essential for signing important matters. By converting content into digital data, the efficiency of data retrieval is significantly improved.

In recent years, machine learning technology has been widely applied across various fields, including handwritten character recognition. Traditional machine learning methods such as Support Vector Machine (SVM), Random Forest (RF), K Nearest Neighbor (KNN), and Decision Tree (DT) have achieved promising results in early handwritten character recognition tasks. For instance, SVM has shown excellent performance in binary classification problems, while KNN has been widely used for its simplicity and effectiveness. However, these methods often struggle with high-dimensional data or images with significant noise interference.

With the advancement in computational power and the rapid development of GPUs, deep learning methods have gradually become the mainstream technology for handwritten character recognition. In particular, Convolutional Neural Networks (CNN) have been widely adopted in handwritten character recognition tasks due to their outstanding performance in image recognition. Additionally, techniques such as Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) have also found applications in tasks involving sequential data and time-dependent data.

This article aims to explore how ensemble learning methods can further improve the performance of handwritten character recognition. We selected three classic deep-learning models and optimized the final classification results through an ensemble learning strategy. Ensemble learning leverages the complementary strengths of different models, effectively enhancing the accuracy and robustness of recognition through weight adjustments.

The primary contribution of this study is to design a handwritten character recognition system based on Ensemble Learning. By combining the strengths of multiple enhanced deep-learning models, this approach significantly improves classification accuracy and demonstrates its effectiveness on the public EMNIST dataset.

The structure of this paper is as follows. Section 2 reviews and analyzes related research in the field of handwritten character recognition. Section 3 introduces the proposed handwritten character recognition model based on Ensemble Learning. Section 4 presents the experimental design and results, comparing the performance of different models on public datasets to demonstrate the superiority of the proposed method. Section 5 summarizes the main contributions and findings of this study and discusses potential future research directions.

2. RELATED WORK

Handwritten character recognition, as a significant research domain in computer vision and pattern recognition, has long been a focal point in both academia and practical applications. With the rapid advancement of machine learning and deep learning technologies, the accuracy and efficiency of handwritten character recognition have significantly improved. Traditional machine learning

Manuscript received September 9, 2024; revised November 18, 2024; accepted December 1, 2024.

[1*] Professor (corresponding author), International Graduate Institute of Artificial Intelligence, National Yunlin University of Science and Technology, Taiwan, R.O.C. (email: changkc@yuntech.edu.tw)

[2] Master Student, International Graduate Institute of Artificial Intelligence, National Yunlin University of Science and Technology, Taiwan, R.O.C.

methods such as Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) played vital roles in early handwritten character recognition. However, as the scale of data expands and the complexity of problems increases, these methods have gradually shown limitations when handling large-scale datasets and diverse handwritten characters.

De Cao Tran et al. [1] employed SVM classifiers to recognize French accented characters and merged the results, thereby reducing recognition errors and improving processing speed. However, this approach is limited in handling larger or more complex character sets due to insufficient feature representation capability. Gauri Katiyar et al. [2] utilized various feature extraction techniques combined with SVM, demonstrating generalization ability in high-dimensional space, but challenges remain in adapting to different fonts and handwriting styles.

Asish Kumar Ghosh et al. [3] combined HOG features with SVM and Bootstrap Aggregating techniques for optical character recognition of Bengali handwritten characters, noting challenges in handling complex character shapes, particularly with “Matra” structures, which can limit performance in certain scenarios. Premanand Ghadekar et al. [4] proposed a hybrid DWT and DCT feature extraction method with KNN and SVM classifiers. While this method performed well on standard datasets, its stability when handling variations in handwriting needs further validation. Sulaiman Khan et al. [5] employed KNN to recognize handwritten Pashto characters, showcasing its applicability to specific character sets, but generalizability to larger datasets and diverse handwriting remains challenging.

Lavanya K et al. [6] compared methods such as Hoeffding Tree, Decision Tree, and Random Forest on the MNIST dataset, demonstrating varying levels of effectiveness. However, these methods face challenges in maintaining stability and efficient model building when applied to more complex handwritten data.

In the realm of deep learning, Chunpeng Wu et al. [7] proposed a model based on Relaxed Convolutional Neural Networks (R-CNN), which enhanced network expressiveness by adopting a relaxation strategy in convolutional layers. Despite the performance improvements, the increased number of parameters and training complexity led to higher computational costs. Samad Roohi et al. [8] used a single Convolutional Neural Network (SCNN) based on LeNet-5 and an Ensemble Convolutional Neural Network (ECNN) for Persian handwritten character recognition. Although ECNN outperformed SCNN, it also increased the complexity of parameter tuning and real-world application.

Bhagyasree P V et al. [9] utilized Directed Acyclic Graph Convolutional Neural Networks (DAG-CNN) to enhance recognition rates for English handwritten characters, demonstrating advantages in multi-directional feature extraction. However, real-time applications were constrained by the high computational load. Rajib Ghosh et al. [10] and Bruno Stuner et al. [11] used BLSTM for recognizing handwritten characters in Indian and French scripts, showcasing the strengths of recurrent networks for sequential data processing. These methods reduced the need for data annotation but risk overfitting with small datasets.

In recent studies, Sarayut Gonwirat et al. [12] proposed the DeblurGAN-CNN framework, integrating DeblurGAN with CNNs such as DenseNet121 and MobileNetV2 to improve the recognition performance of noisy handwritten characters, particularly in low-contrast and blurred images. Although the accuracy was high, integrating more approaches could further ensure precision. Aqsa Rasheed et al. [13] employed pre-trained AlexNet for Urdu character and digit recognition, incorporating transfer learning and data augmentation techniques to enhance performance. While their approach showed promising results across multiple datasets, incorporating different lightweight CNN architectures could improve recognition accuracy and stability.

In summary, existing machine learning methods have shown contributions to small-scale and specific-language handwritten character recognition. However, as data scale expands and character diversity increases, their limitations become more apparent. Deep learning techniques, particularly Convolutional Neural Networks, have demonstrated significant advantages in improving recognition performance, but challenges remain in addressing data heterogeneity and high computational resource requirements. Therefore, this study proposes an ensemble learning approach that leverages the strengths of multiple deep learning models to enhance the accuracy and robustness of handwritten character recognition, addressing the shortcomings of existing technologies in handling diverse and complex data.

3. PROPOSED METHOD

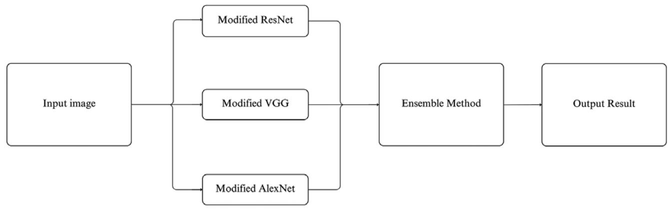


Fig. 1 The Architecture of the Proposed Ensemble Method.

In the aforementioned research, we observed that while machine learning and deep learning techniques have made significant advancements in handwritten character recognition, the performance of individual models often remains limited across different datasets. This limitation is particularly pronounced when handling diverse handwritten characters, where there is still room for improvement in robustness and accuracy. Numerous studies have highlighted that future work should focus on integrating multiple CNN model frameworks to achieve higher recognition accuracy and stability. To address this challenge, this article proposes a handwritten character recognition method based on ensemble learning. Figure 1 illustrates the architecture of the proposed method; by leveraging the strengths of multiple deep learning models, we designed an ensemble strategy to enhance recognition accuracy.

Specifically, we selected three well-known lightweight deep learning architectures—VGG, ResNet, and AlexNet—as base models. These models have demonstrated strong performance on classic handwritten character datasets, and we made targeted adjustments to better align with the needs of our research. In this ensemble approach, each base model makes independent predictions, which are subsequently combined using a weighted averaging method. This strategy not only yielded promising results for individual models but also demonstrated competitive performance on the public EMNIST dataset.

3.1 Enhanced AlexNet

Layer Name	Layer Details	Output Size
Input	_	(28, 28, 3)
Conv 1	5x5, ReLU	(24, 24, 96)
Maxpool 1	3x3 window, stride 2	(11, 11, 96)
Conv 2	5x5, ReLU	(11, 11, 256)
Maxpool 2	3x3 window, stride 2	(5, 5, 256)
Conv 3	3x3, ReLU	(5, 5, 384)
Conv 4	3x3, ReLU	(5, 5, 384)
Conv 5	3x3, ReLU	(5, 5, 256)
Maxpool 2	3x3 window, stride 2	(2, 2, 256)
Flatten	_	(1024)
Dense 1	4096, ReLu	(4096)
Dense 2	4096, ReLu	(4096)
Output	16, SoftMax	(16)

Fig. 2 The Model Structure of Modified AlexNet.

AlexNet, a milestone model in the field of deep learning, established its significance in Convolutional Neural Networks (CNN) with its outstanding performance in the 2012 ImageNet competition. Due to its relatively simple architecture, AlexNet uses fewer parameters and layers, making it efficient in terms of training speed and resource consumption. For the task of handwritten character recognition, AlexNet's features make it an ideal choice, especially when dealing with smaller datasets, as it can quickly extract effective features and perform efficient classification.

The original architecture of AlexNet consists of five convolutional layers, each followed by a ReLU activation function and a max-pooling layer. These convolutional layers are designed to extract features at different levels, which are then fused and classified through three fully connected layers. As shown in Figure 2, to adapt to the specific requirements of handwritten character recognition, we adjusted the input layer to fit the image size of the dataset and set the number of nodes in the output layer to match the number of classes for more accurate classification.

3.2 Enhanced VGG

Layer Name	Layer Details	Output Size
Input	_	(28, 28, 3)
Conv 1	3x3, ReLU	(28, 28, 64)
Conv 2	3x3, ReLU	(28, 28, 64)
Maxpool 1	3x3 window	(14, 14, 64)
Conv 3	3x3, ReLU	(14, 14, 128)
Conv 4	3x3, ReLU	(14, 14, 128)
Maxpool 2	2x2 window, stride 2	(7, 7, 128)
Conv 5	3x3, ReLU	(7, 7, 256)
Conv 6	3x3, ReLU	(7, 7, 256)
Conv 7	3x3, ReLU	(7, 7, 256)
Maxpool 3	2x2 window, stride 2	(3, 3, 256)
Flatten	_	(2304)
Dense 1	4096, ReLu	(4096)
Dense 2	4096, ReLu	(4096)
Output	16, SoftMax	(16)

Fig. 3 The Model Structure of Modified VGG.

The design of the VGG model is based on a deep stack of multiple convolutional layers, each followed by a ReLU activation function to enhance the model's nonlinear representation capabilities. These convolutional layers are designed to progressively extract different levels of features from images, ranging from edges to higher-level shapes and patterns. As shown in Figure 3, to accommodate the 28x28 pixel input of handwritten character images, we adjusted the VGG model by reducing the number of convolutional layers. This adjustment preserves the feature extraction capabilities while avoiding excessive computational costs.

3.3 Enhanced ResNet

ResNet (Residual Network) represents a significant breakthrough in deep learning, effectively addressing the common issue of vanishing gradients in deep networks through its residual block design. Proposed by He et al. in 2015, the ResNet model stood out in the ImageNet competition and quickly became a benchmark in deep learning. The residual structure of ResNet allows signals to bypass layers, enabling the training of very deep networks while maintaining efficient feature extraction capabilities.

Layer Name	Layer Details	Output Size
Input	_	(28, 28, 3)
Conv 1	3x3, BN, ReLU	(28, 28, 64)
Maxpool 1	3x3 window, stride 2	(14, 14, 64)
Residual Block 1	(3x3, BN, ReLU) x2	(14, 14, 64)
Residual Block 2	(3x3, BN, ReLU) x2	(14, 14, 64)
Residual Block 3	(3x3, BN, ReLU) x2	(14, 14, 64)
GlobalAvgPool	Global Average Pooling	(64)
Flatten	_	(64)
Dense 1	256, ReLU	(256)
Dense 2	128, ReLU	(128)
Output	16, SoftMax	(16)

Fig. 4 The Model Structure of Modified ResNet.

As shown in Figure 4, the core feature of ResNet is its unique residual block design, where each block contains two convolutional layers with a skip connection between them. This skip connection allows the input signal to be directly passed to the output of the block, reducing the problem of vanishing gradients in deep networks. This structure allows ResNet to train very deep networks while preserving strong learning capabilities.

In our study, we selected and adjusted three renowned deep learning models—AlexNet, VGG, and ResNet—to design a robust and efficient base model ensemble for the handwritten character recognition task. Each model's architecture was specifically tailored to fit the dataset format, and data augmentation techniques and optimized training strategies were employed to enhance the models' generalization abilities. Building on this foundation, we further improved recognition performance by combining these models into an ensemble system, leveraging the strengths of each to achieve higher recognition accuracy.

3.4 Ensemble Strategy

In this study, we adopted a weighted averaging ensemble strategy to combine the prediction results of the three base models: AlexNet, VGG, and ResNet. The purpose of this strategy is to maximize the overall prediction accuracy by reasonably distributing the weights of each model. The weighted averaging strategy not only takes advantage of the characteristics of each model but also reduces the errors of individual models to some extent, thereby improving the final classification performance.

To implement the weighted averaging ensemble, we set the weights of the three models, w1, w2, and w3, such that their sum is 1, i.e., w1 + w2 + w3 = 1, and each weight value ranges from 0 to 1 with a step size of 0.1. This results in a total of 66 possible weight combinations, which we tested individually to determine the optimal weight distribution. The weight adjustment process can be represented by the Equation (1) :

$$W = \left\{ \begin{matrix} (w1, w2, w3) \ | \ w1, w2, w3 \in \{0.1, 0.2, \ldots, 0.8\}, \\ w1 + w2 + w3 = 1 \end{matrix} \right\} \tag{1}$$

The final ensemble prediction result is calculated using the Equation (2) :

$$\hat{y} = w1 \cdot y\hat{1} + w2 \cdot y\hat{2} + w3 \cdot y\hat{3} \tag{2}$$

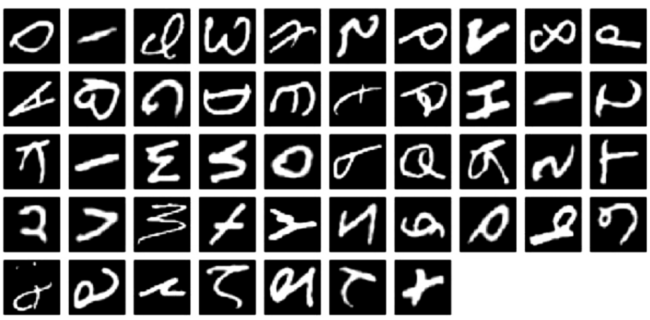
Where $\hat{y}$ is the final prediction result of the ensemble strategy, and $y\hat{1}$, $y\hat{2}$, and $y\hat{3}$ are the prediction results of AlexNet, VGG, and ResNet, respectively. This formula shows that by using the weighted averaging method, we combine the prediction results of the three models to improve prediction accuracy.

By combining the strengths of AlexNet, VGG, and ResNet through the weighted averaging ensemble strategy, we effectively enhanced the accuracy of handwritten character recognition in high-noise environments. This strategy not only improved the overall stability of the predictions but also increased the accuracy, surpassing the performance of any single model.

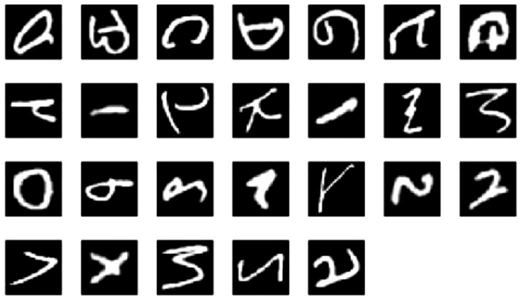
4. EXPERIMENTAL RESULT

4.1 Dataset

The EMNIST (Extended MNIST) dataset [14] is an extension of the classic MNIST dataset, designed to provide a more challenging benchmark for handwritten character recognition. Proposed by Cohen et al. in 2017, the EMNIST dataset is derived from the NIST Special Database 19, which contains a large number of handwritten digit and letter samples. The EMNIST dataset retains the basic structure and image format of the MNIST dataset, making it directly compatible with existing MNIST classifiers while offering a larger dataset with more classes, covering both handwritten letters and digits. As a result, the EMNIST dataset is suitable not only for handwritten digit classification but also for handwritten letter classification, providing researchers with a more comprehensive testing platform.



(a) EMNIST-balanced dataset



(b) EMNIST-letters dataset

Fig. 5 Examples from the publicly available EMNIST handwritten character dataset. (a) represents the Balanced subset, and (b) represents the Letters subset.

The EMNIST dataset consists of five subsets, each designed for different classification tasks:

- ByClass: Contains 62 classes (10 digits and 52 letters), including both uppercase and lowercase letters, with a total of 814,255 samples.
- ByMerge: Merges the uppercase and lowercase forms of letters, reducing the letters to 37 classes. Including the digit classes, there are 47 classes in total, with 814,255 samples.
- Balanced: Provides a balanced subset of letters and digits with 47 classes. There are 112,800 training samples and 18,800 test samples, totaling 131,600 samples.
- Digits: A dataset of pure digits containing the 10-digit classes from MNIST, with a total of 280,000 samples.
- Letters: A dataset containing only letters, with 26 classes. There are 88,800 training samples and 14,800 test samples, totaling 103,600 samples.

Each subset is divided into training and testing sets, with sample images being 28x28 pixel grayscale images. These samples have undergone standardized preprocessing, including Gaussian blurring and resampling, to ensure compatibility with the original MNIST dataset. This allows the EMNIST dataset to support both digit and letter recognition tasks, providing data support for various application scenarios. As shown in Figure 5, to avoid recognition bias caused by class imbalance, we used the "Balanced" and "Letters" subsets of the EMNIST dataset as benchmark datasets to validate the competitiveness of the proposed method on public datasets, further demonstrating its broad applicability in handwritten character recognition. These subsets present different challenges and allow us to evaluate the proposed Ensemble method across various scenarios.

4.2 Training Results of Individual Models

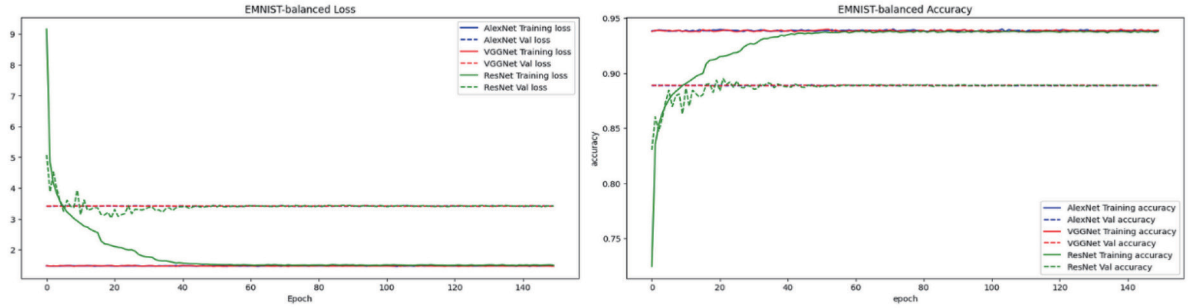


Fig. 6 The Loss & Accuracy during Training Process based on EMNIST-Balanced subset.

To validate the applicability of our method on public datasets, we selected the EMNIST dataset for experimentation and compared the results with existing studies. First, we retrained three base models (Modified ResNet, Modified AlexNet, Modified VGG) on the EMNIST dataset. After testing, we chose the Adam optimizer and categorical_crossentropy loss function and determined the optimal learning rate and batch size for each model through grid search. During training, we applied various data augmentation techniques, such as rotation, translation, and scaling, to effectively enhance the models' generalization ability. To prevent overfitting, we introduced EarlyStopping and ReduceLROnPlateau strategies, ensuring that the models achieved optimal performance while converging stably.

Figure 6 shows the Loss & Accuracy during the training process on the EMNIST-balanced dataset, the three models exhibited different convergence behaviors. The loss for AlexNet and VGG reached relatively low levels early in the training process and remained stable, with no significant changes thereafter. In contrast, ResNet's loss started higher but gradually decreased as training progressed, eventually reaching a level comparable to that of AlexNet and VGG. This indicates that ResNet was able to overcome its initial larger errors and eventually stabilized. Correspondingly, the accuracy of AlexNet and VGG was already at a high level early in the training and remained stable throughout. ResNet, on the other hand, showed a gradual improvement in accuracy as training progressed, ultimately reaching a level similar to the other two models, demonstrating its potential in incremental learning.

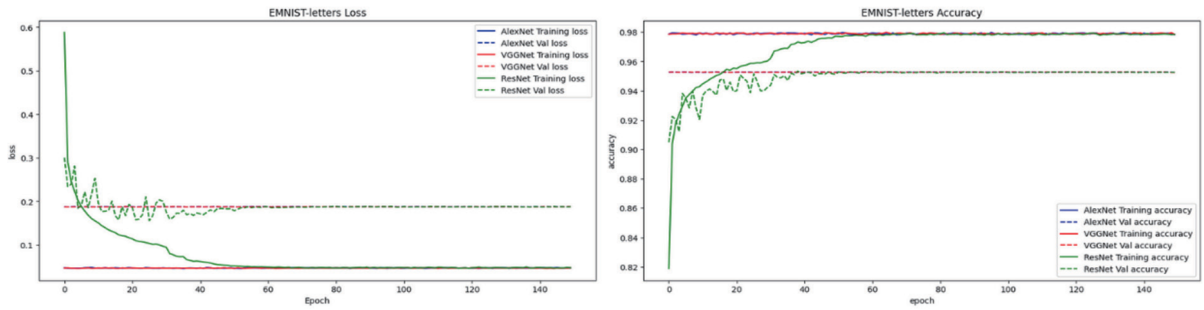


Fig. 7 The Loss & Accuracy during Training Process based on EMNIST-Letters subset.

Figure 7 shows the Loss and Accuracy results during the training process on the EMNIST-letters dataset, which showed a similar trend. The loss for VGG and AlexNet started low and remained stable without significant changes during subsequent training. ResNet, again, showed a process of gradual optimization, with initially higher loss that gradually decreased, eventually reaching a level similar to VGG and AlexNet. Similarly, in terms of accuracy, VGG and AlexNet achieved high accuracy levels early in the training, which remained stable throughout. ResNet gradually improved its accuracy during training, eventually reaching a level comparable to VGG and AlexNet. Overall, these results indicate that ResNet demonstrated strong adaptability during training, improving its performance incrementally, while VGG and AlexNet showed stable and high performance from the beginning.

4.3 Experimental Results of Ensemble Strategy

Figure 8 shows the performance of the ensemble strategy based on different weight combinations on the EMNIST-balanced and EMNIST-letters datasets. In the EMNIST-balanced dataset, when the weight combination is 0.3, 0.2, 0.5 (i.e., ResNet, AlexNet, and VGG account for 30%, 20%, and 50% of the weight, respectively), the model achieves the highest accuracy of approximately 0.9064. This suggests that giving a higher weight to the VGG model can more effectively enhance overall recognition performance. In the EMNIST-letters dataset, the highest accuracy of approximately 0.9456 is achieved with a weight combination of 0.3, 0.4, 0.3, indicating that a balanced distribution of weights among the three models, particularly with more weight given to AlexNet, can maximize performance on the letters subset. These results highlight the importance of adjusting weight strategies based on the specific dataset in ensemble learning to optimize final prediction accuracy.

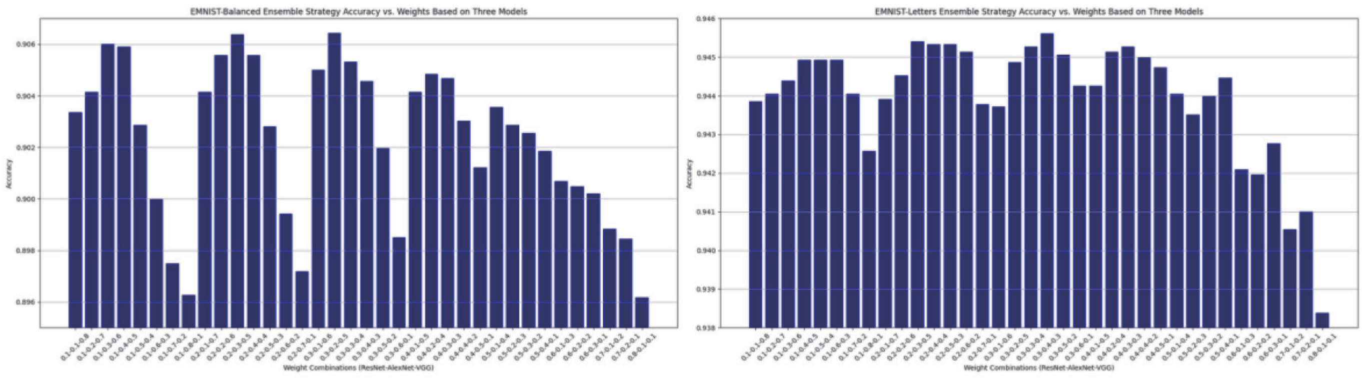


Fig. 8 The Ensemble results with different weight combinations. (Left) is based on the EMNIST-Balanced subset, and (Right) is based on the EMNIST-Letters subset.

Figure 9 shows the analysis of the Confusion Matrix based on the EMNIST-balanced and EMNIST-letters datasets, which shows that the ensemble model significantly outperforms individual models in prediction accuracy. Particularly for some difficult-to-distinguish classes, the ensemble strategy successfully reduces the error rate, demonstrating an overall performance improvement. For the EMNIST-balanced dataset, the ensemble model's predictions are more precise, with a noticeable reduction in misclassifications. Similarly, as shown in Figure 10, in the EMNIST-letters dataset, the ensemble model performs excellently, effectively reducing prediction errors for certain letters made by individual models. Additionally, since the test dataset only includes the first 19 classes, the unpredicted classes reflect the dataset's limitations rather than a flaw in the model. These results further confirm the effectiveness of the ensemble strategy in handwritten character recognition, particularly in handling challenging datasets, where the ensemble method can significantly improve the model's prediction accuracy.

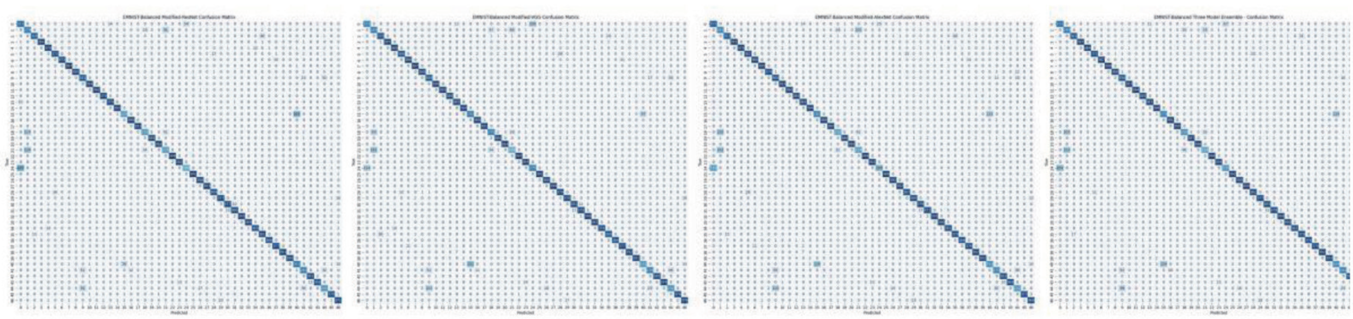


Fig. 9 The Confusion Matrix of Three Single Model and Ensemble Method based on EMNIST-Balanced subset. (From left to right are Modified ResNet, VGG, AlexNet, and the Ensemble Method.)

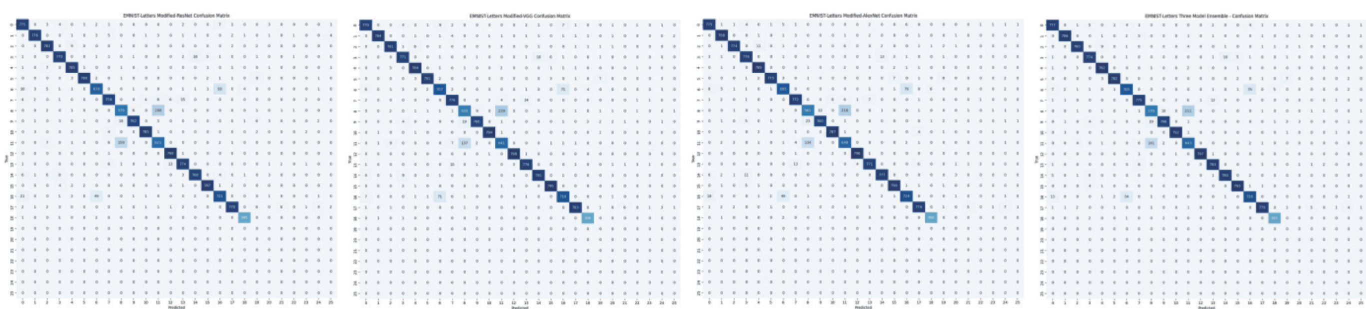


Fig. 10 The Confusion Matrix of Three Single Model and Ensemble Method based on EMNIST-Letters subset. (From left to right are Modified ResNet, VGG, AlexNet, and the Ensemble Method.)

4.4 Ablation Experiment

Table 1 shows the results of ablation experiments; we compared the performance of individual models, pairwise combinations, and the three-model ensemble. The individual models (ResNet, AlexNet, VGG) performed well on both datasets, but their accuracy was slightly lower than some more complex deep learning architectures. When we combined two models, the accuracy improved, especially the VGG + AlexNet combination, which achieved 90.51% on the balanced dataset, approaching the performance of more complex models. When the three models were combined using weighted averaging, the accuracy further increased, reaching 90.64% on the balanced dataset and 94.56% on the letters dataset. These results indicate that the ensemble strategy effectively leveraged the strengths of different models, significantly improving the accuracy of handwritten character recognition and achieving the best performance on the EMNIST-balanced dataset.

Table 1 Comparison of the results from the Ablation Experiment and other studies.

Technique	Balanced (chars + letters)	Letters
Linear classifier [14]	50.93%	55.78%
OPIUM [14]	78.02%	85.15%
CNN(6conv+2dense) [15]	90.59%	-
Markov random field CNN [16]	90.29%	95.44%
Text Caps [17]	90.46%	95.36%
CNN(flat, 2conv+1dense) [18]	87.18%	93.63%
ResNet	89.37%	93.68%
AlexNet	89.38%	94.30%
VGG	90.05%	94.13%
ResNet+AlexNet	89.96%	94.39%
ResNet+VGG	90.39%	94.37%
VGG+AlexNet	90.51%	94.45%
Ensemble Method (ours)	90.64%	94.56%

The ResNet, AlexNet, and VGG models mentioned in the table are Modified versions with architecture adjustments made to fit this study.

In this section, we conducted detailed experiments and analysis of the three deep learning models (ResNet, AlexNet, VGG) based on the public EMNIST dataset. The individual models exhibited different convergence behaviors during training, with the ResNet model gradually catching up to the other two models in the early stages of training. The experimental results of the ensemble strategy showed that the weighted averaging strategy effectively enhanced the overall prediction accuracy of the models, especially on more challenging datasets, where the ensemble model outperformed the individual models. Additionally, the ablation experiments further demonstrated the potential of ensemble learning in combining the strengths of different models, effectively improving the stability and accuracy of handwritten character recognition.

4.5 Grad-CAM Visualization and Analysis of Ensemble Models

Figure 11 presents the model visualization heatmaps generated via Grad-CAM. From these visualizations, it can be observed that each model exhibits distinct attention distributions over the input images. This diversity in attention is a key advantage of the ensemble approach: by leveraging the strengths of different models, the overall predictive performance and stability can be significantly improved.

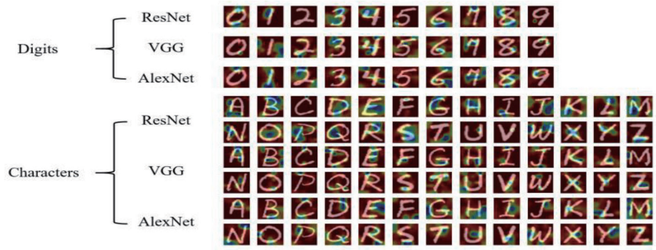


Fig. 11 Heatmap Visualization of Model Attention Patterns based on EMNIST-balanced (Top - digits, Bottom - characters).

However, an interesting phenomenon can also be seen in the heatmaps, where the red regions cover most parts of the image, including the black background. This suggests that the model may not have fully learned to disregard the influence of the background during training, instead incorporating it into the decision-making process. Such behavior indicates that the model might develop misleading dependencies on irrelevant features, which is another area where ensemble methods can bring improvements. Individual models may introduce biases toward certain irrelevant features, but by combining multiple models, the impact of these biases can be mitigated, leading to more accurate and reliable predictions.

Moreover, it is noteworthy that the three base models we employed (ResNet, VGG, AlexNet) are all relatively lightweight architectures. Without extensive optimization, these models may struggle to achieve optimal performance when dealing with the complexity of the dataset. In this context, the ensemble strategy has proven to be an effective and straightforward means of significantly enhancing prediction accuracy and stability. While each model on its own might not fully distinguish the importance between “digit” and “background” features, the ensemble strategy helps to compensate for these individual shortcomings, thereby improving overall model performance.

These findings clearly demonstrate the superiority of ensemble learning when dealing with diverse model characteristics. By aggregating models with different architectures, we can harness their respective strengths, resulting in a more robust system that reduces dependency on specific regions or features and thereby enhances the generalization capability of the model. This conclusion is also insightful for future model design and optimization, especially when balancing performance and computational efficiency. Ensemble methods offer a practical and feasible approach for achieving such improvements.

5. CONCLUSION

This article proposes a handwritten character recognition method based on ensemble learning, which improves accuracy by combining the strengths of ResNet, AlexNet, and VGG deep learning models. The experimental results demonstrate that this ensemble strategy performs excellently on the public EMNIST dataset, not only surpassing the predictive performance of individual models but also showing greater robustness and generalization ability in handling the challenges of different types of datasets. Future work will continue to explore ways to further optimize the ensemble strategy and apply this method to a broader range of handwritten character recognition scenarios to enhance its effectiveness in practical applications.

REFERENCES

[1] De Cao Tran, P.F., Jean-Marc Ogier, *Accented Handwritten Character Recognition Using SVM - Application to French,* in 2010 12th International Conference on Frontiers in Handwriting Recognition. 2010. p. 65-71.

[2] Gauri Katiyar, S.M., *SVM based off-line handwritten digit recognition, in 2015 Annual IEEE India Conference (INDICON).* 2015.

[3] Asish Kumar Ghosh, S.A., *A comparison between Support Vector Machine SVM and bootstrap aggregating technique for recognizing Bangla handwritten characters, in 2017 20th International Conference of Computer and Information Technology (ICCIT).* 2017.

[4] Premanand Ghadekar, S.I., Dhruv Sonone, *Handwritten Digit and Letter Recognition Using Hybrid DWT-DCT with KNN and SVM Classifier, in 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA).* 2018 p. 1-6.

[5] Sulaiman Khan, H.U.K., *Isolated Handwritten Pashto Characters Recognition using KNN classifier, in 2021 International Conference on Electrical, Computer and Energy Technologies (ICECET).* 2021. p. 1-5.

[6] Lavanya K, S.B., Prateek Tank, Shashwat Jain, *Handwritten digit recognition using hoeffding tree, decision tree and random forests — A comparative approach, in 2017 International Conference on Computational Intelligence in Data Science(ICCIDS).* 2017 p. 1-6.

[7] Chunpeng Wu, W.F., Yuan He, Jun Sun, Satoshi Naoi, *Handwritten Character Recognition by Alternately Trained Relaxation Convolutional Neural Network, in 2014 14th International Conference on Frontiers in Handwriting Recognition.* 2014. p. 291-296.

[8] Samad Roohi, B.A., *Persian handwritten character recognition using convolutional neural network, in 2017 10th Iranian Conference on Machine Vision and Image Processing (MVIP).* 2017. p. 247-251.

[9] Bhagyasree P V, A.J., Chandran Saravanan, *A Proposed Framework for Recognition of Handwritten Cursive English Characters using DAG-CNN, in 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT).* 2019 p. 1-4.

[10] Rajib Ghosh, P.K., Prabhat Kumar, *RNN Based Online Handwritten Word Recognition in Devanagari Script, in 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR).* 2018. p. 517-522.

[11] Bruno Stuner, C.e.C., Thierry Paquet, *Self-Training of BLSTM with Lexicon Verification for Handwriting Recognition, in 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR).* 2017. p. 633-638.

[12] Sarayut Gonwirat, O.S., *DeblurGAN-CNN: Effective Image Denoising and Recognition for Noisy Handwritten Characters.* IEEE Access, 2022. 10: p. 90133-90148.

[13] Aqsa Rasheed, N.A., Bushrazafar, Amsashabbir, Muhammadsajid, Muhammadtariqmahmood, *Handwritten Urdu Characters and Digits Recognition Using Transfer Learning and Augmentation With AlexNet.* IEEE Access, 2022. 10: p. 102629-102645.

[14] Gregory Cohen, S.A., Jonathan Tapson, Andr´e van Schaik, *EMNIST: an extension of MNIST to handwritten letters.* ArXiv abs/1702.05373, 2017.

[15] A. Shawon, M.J.-U.R., F. Mahmud, M. M. Arefin Zaman, *Bangla Handwritten Digit Recognition Using Deep CNN for Large and Unbiased Dataset, in 2018 International Conference on Bangla Speech and Language Processing (ICBSLP).* 2018 p. 1-6.

[16] Peng, Y.a.Y., Hujun, *Markov Random Field Based Convolutional Neural Networks for Image Classification.* 2017: p. 387-396.

[17] Sabour, S., Nicholas Frosst, Geoffrey E. Hinton, *Dynamic routing between capsules, in Advances in neural information processing systems* **30** (2017). 2017.

[18] Cavalin, P., Soares de Oliveira, Luiz, *Confusion Matrix-Based Building of Hierarchical Classification, in 23rd Iberoamerican Congress, CIARP 2018, Madrid, Spain.* 2019. p. 271-278.