American Journal of Biochemistry and Biotechnology

Original Research Paper

Capsule Defect Detection and Classification Using Enhanced MobileNet Model

[1]Wenqing Bian, [1]Yongzhi Huang, [1]Baolian Shan and [2]Yanting Chang

[1]*Department of Intelligent Medical Engineering, Academy of Medical Engineering and Translational Medicine,*
Tianjin University, 92 Weijin Road, Weijin Road Campus, Nankai District, Tianjin, China
[2]*Tianjin Changqing Technology Development Co., Ltd., No. 7, Changtu Road, Hongqiao District, Tianjin, China*

Article history
Received: 17-05-2024
Revised: 03-07-2024
Accepted: 09-08-2024

Corresponding Author:
Yongzhi Huang
Academy of Medical Engineering and Translational Medicine, Tianjin University, Tianjin, China
Email: yongzhi_huang@tju.edu.cn

Abstract: The quality of capsules has a direct impact on human health. However, defects inevitably arise during the capsule manufacturing process, and manual sorting is required. Most studies on capsule defects focus on detecting surface imperfections and those on internal capsule defects remain relatively limited. This study classifies four types of capsules using the deep learning model MobileNet, aiming to accurately identify surface and internal defects. A total of 2872 capsule images are used to evaluate the model's classification performance prior to and following optimization. The MobileNet model categorizes capsule images into four types: Normal capsules, deformed capsules, impure capsules, and bubbled capsules. Stratified cross-validation is applied to partition the dataset into 80% training, 10% validation, and 10% testing sets for ten-fold cross-validation. The model's performance is evaluated using three metrics: Precision, recall, and F1-score. The results are compared with two classic deep learning models, four traditional machine learning models (VGG16, ResNet101, KNN, and SVM), as well as decision trees and random forests. The findings demonstrate that the MobileNet model exceeds the performance of the other models, achieving precision, recall, and F1 scores of 94.24, 94.75 and 94.23%, respectively. Through transfer learning and improving the top layers of the MobileNet model with dropout, L2 regularization, Batch Normalization (BN), and average pooling, the model accuracy is improved by 7.95%, indicating promising performance and potential in detecting capsule defects.

Keywords: MobileNet, Capsule Defects, Image Processing, Deep Learning, Transfer Learning

Introduction

The quality of pharmaceuticals is crucial in the medical industry and has a direct impact on human health. Capsule-based medications are of great importance in healthcare, but they often suffer from various defects due to limitations in production technology. These defects, ranging from air bubbles to shrinkage and deformities, result in inconsistent dosages, compromised efficacy, and poor sealing, which will affect the potency and shelf life of medication. Cosmetic defects have little effect on the efficacy of medications, but they can reduce the image of consumers. In the last decade, due to the lack of systematic equipment for detecting capsule defects, manual visual inspection and sampling have often been used to detect capsule defects. However, these approaches have low detection rates, limited detection accuracy, and susceptibility to individual subjective factors (Zhou *et al*., 2020).

Actually, capsule defect detection can be summarized as surface defects and internal defects detection. Defect detection is an image classification and computer vision problem. Currently, computer vision-based online capsule detection technologies are often utilized in the medical industry. Based on the traditional image segmentation methods, a scanning-based area growth technique was adopted to achieve online capsule defect detection (Zhu *et al*., 2011). The RGB color space can establish an algebraic relationship between the three components for the segmentation of capsule defects, which is useful for detecting capsule defects and reducing the processing time in real-time capsule defect identification systems (Hou *et al*., 2011).

The rapid rise and development of artificial intelligence applications have led to significant growth in intelligent image-processing devices. Convolutional Neural Networks (CNN) and other deep learning techniques are employed for

Science Publications

classifying images (Pławiak and Acharya, 2020). CNN-based Inception-ResNet-V2 model was used for the assessment of damage in wooden structures (Ehtisham *et al.*, 2024). The CNN-based MobileNet model was used to automatically recognize the position of the mask on the face, achieving high accuracy and F1-score (Rahman *et al.*, 2023). A novel vision transformer-based deep neural network is proposed by integrating the transformer with CNN for image classification.

The advanced CNN model MobileNet is frequently used in medical diagnostic tasks and agriculture, particularly for the classification of skin diseases, assessment of lung diseases by chest radiographs (Heidari *et al.*, 2020), identification of implanted ridges on CBCT slices for classification (Chang *et al.*, 2024); YVMV-infected okra plants (Chawla *et al.*, 2024) and automatic classification of plant diseases (Sutaji and Yıldız, 2022). Unlike traditional CNN architectures, MobileNet employs depthwise separable convolutions to independently filter input and combine outputs in channels. This innovative approach can significantly reduce computation and overall model complexity. Its lightweight architecture can be efficiently implemented on devices with limited resources, such as mobile devices and embedded platforms, which is particularly beneficial for dental practices with constrained high-performance computing power (Mohapatra *et al.*, 2021). In recognition of novel coronavirus, the classification of MobileNetV2-Support Vector Machine is optimal (Surono *et al.*, 2023a).

Through thorough image preprocessing, MobileNet has achieved performance comparable to that of traditional CNNs, highlighting its potential for widespread adoption in medical diagnostics (Howard *et al.*, 2017). The appearance of capsules has different characteristics. Previous studies did not use the MobileNet algorithm for capsule defect detection. Usually, existing studies on capsule defect detection mainly focus on surface defects and a few on internal capsule defects. First, a machine vision system is designed to comprehensively capture all surfaces of the capsules, which involves a stable mechanical structure and poses considerable challenges. Second, due to the wide variety and often subtle nature of defects, their identification is particularly complex. The capsules have a diameter of 6 mm and the minimum detectable defect size is 0.1 mm, presenting significant technological challenges for research. Internally, capsules are filled with gel. When the smart camera shoots, the impurities and bubbles inside can reduce image clarity. Specifically, when there are no defects on the front of the capsule, but on the back, the captured defects are even more blurred, which poses a significant challenge to model learning. Therefore, this study aims to assess the application potential of MobileNet in detecting defects on the back and inside of capsules.

Materials and Methods

Design of Model

The MobileNet structure is shown in Fig. (1). It includes two types of convolution layers (depthwise and pointwise), along with pooling layers, and is followed by fully connected hidden layers and a Softmax layer. The convolution layer uses a set of learnable filters for the convolution of the input image. Each convolutional kernel is represented as a compact weight matrix that slides over the input image, computing dot products at every position. This method produces a collection of feature maps that emphasize the unique attributes of the input images. The depthwise convolution kernel layer is 3×3 and the pointwise convolution layer is 1×1×3. Maximum pooling reduces dimensions, while ReLU introduces non-linearity for better learning. The Softmax layer receives a numerical vector as input and generates a probability map over a group of separate, non-intersecting categories. The Softmax function enhances the separability between different classes by increasing the distance between them in the transformed feature space (Akil *et al.*, 2020). The values of each input vector indicate the level of evidence for each category. The Softmax function (Eq. 1) converts these values into probabilities that sum to 1 and typically selects the category with the highest probability as the prediction. The optimization process for each model uses an image enhancement and transfer learning technique (Naranjo-Torres *et al.*, 2020): Fine-tuning (finetune).

Table (1) displays the parameters for the MobileNet model. The model is based on the MobileNet architecture, with the top layer removed and the dropout layer added, the BN layer, the average pooling layer, and L2 regularization. Dropout removes random connections between nodes of both layers (Nguyen *et al.*, 2021). The role of dropout and L2 is to prevent overfitting. The proposed MobileNet architecture can be trained for up to 100 consecutive epochs with a batch size of 64. Cross-entropy is utilized as the loss function (Eq. 2) and it rises as the predicted probability diverges from the true label (Chawla *et al.*, 2024). The model seeks to reduce cross-entropy. Adam initialized with a learning rate of 0.01, is utilized to speed up model training and enhance convergence. Model Checkpoint is applied to callbacks to monitor the model's validation loss. It only saves the model if the loss improves relative to the previously best-performing model. To improve the computational efficiency, the proposed model is implemented using Graphics Processing Unit (GPU). Tensor Flow toolbox is used to implement the model:

$$p_j = \frac{e^{x_j}}{\sum_{k=1}^{K} e^{x_K}} \tag{1}$$

$$CE(\theta) = -\sum_{j=1}^{n} y_j \cdot \log(P_j) \tag{2}$$

where, p_j is the j^{th} component of the predicted vector of class j;j is an integer with a range from 1 to K; K is the number of classes;$CE(\theta)$ is a loss function that represents the error between the estimated vector andy_j is the j^{th} component of the normalized reference vector.

Design of Experiments and Image Processing

The proposed methodology involves two primary phases: Categorizing capsule defects and calculating their attributes. These phases are illustrated in Fig. (2). The dataset is collected and segmented into four categories: Normal, foreign, impurity, and bubble. Following image pre-processing, a pre-trained MobileNet model is employed for training and validation with this dataset.

To optimize storage space and computational load, the color image should be first converted to grayscale. After that, Gaussian filters are used to smooth the image. These filters can be convolved using the Gaussian function. Last but not least, applying histogram equalization can improve the image contrast.

Table 1: Parameters of the MobileNet model

Layer	Output shape	Param #
Inception_input	(224, 224, 3)	0
Mobilenet_1.00_224	/	3228864
Dropout	(7, 7, 1024)	0
Batch_normalization	(7, 7, 1024)	4096
L2 regularization	/	/
Global_average_pooling2d	1024	0
Dense	512	528900
Dense_1	4	2068

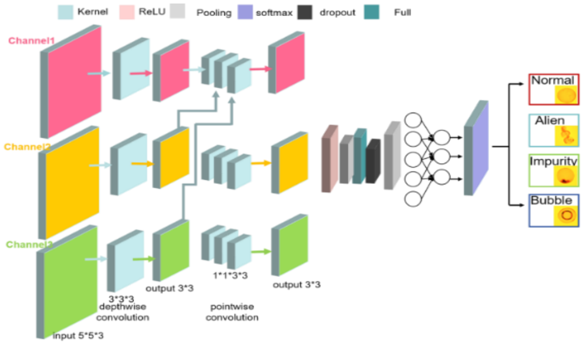


Fig. 1: Architecture of the proposed MobileNet

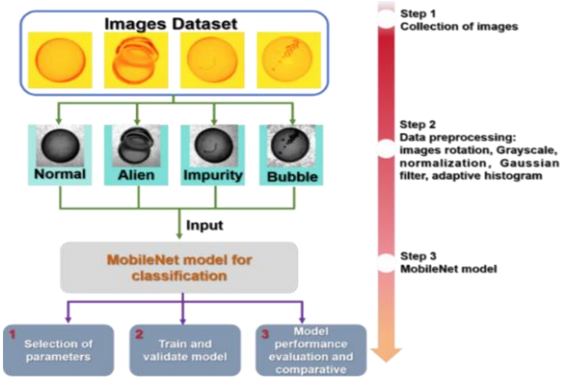
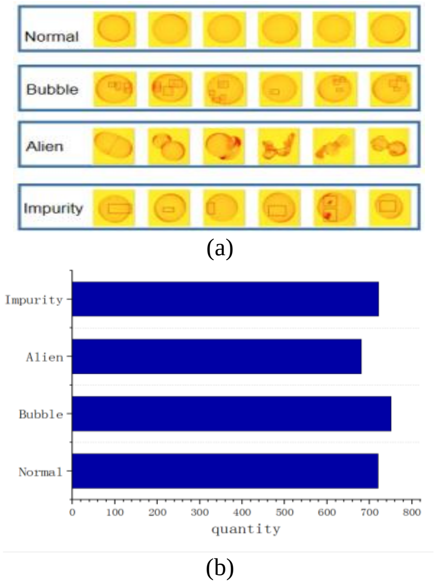

Images Dataset
Normal
Alien
Impurity
Bubble
Input
MobileNet model for classification
1 Selection of parameters
2 Train and validate model
3 Model performance evaluation and comparative
Step 1
Collection of images
Step 2
Data preprocessing:
images rotation, Grayscale,
normalization, Gaussian
filter, adaptive histogram
Step 3
MobileNet model

Fig. 2: Flowchart of the proposed MobileNet methodology

Data Collection and Preparation

During the sorting process following capsule production, four types of images were categorized: Three representing defective capsules and one depicting a qualified capsule (Fig. 3a). The four types of capsules are Normal, alien, impurity, and bubble capsules. In the field of computer vision, the quality standards for images captured by different cameras vary and low-resolution images present a significant challenge for image processing tasks (Surono *et al*., 2023b). All images were captured using an industrial camera (MV-CS05013-10GC, HIKVISION) with a resolution of 2448×2048 pixels and 12 megapixels, 2872 color images in all. 2296 of them are used for training, 288 for validating, and 288 for testing. Considering the memory consumption, the collected capsule images are resized into a resolution of 224×224, and stored with the corresponding image labels as the input dataset. The quality of capsules, from good to poor, is normal capsules (720 images), alien capsules (751 images), impurity capsules (680 images), and bubble capsules (721 images), respectively. The four types of capsule data are largely balanced (Fig. 3b) to reduce the risk of overfitting and accuracy bias. The minimum diameter of bubbles in the bubble capsules reaches 0.1 mm.

Image rotation is used to increase diversity and avoid overfitting in the dataset. Data augmentation (Zhou *et al*., 2022) through width and height shifting (10%), zooming (20%) and rotation (20%) is carried out. Grayscale processing, Gaussian filter, and normalization are also carried out. K-fold cross-validation (K = 10) is performed to assess the training consistency across the entire dataset. The implementation process involved utilizing the K-fold function within the sklern. model_selection module in the Scikit-learn library.



(a)

(b)

Fig. 3: Capsule classification (a) Capsule legend (b) Quantity distribution

Model Performance Evaluation

True Positive" (TP) represents correct predictions for positive instances, "True Negative" (TN) for negative instances, "False Positive" (FP) for incorrect predictions of positive instances, and "False Negative" (FN) for incorrect predictions of negative instances (Thanathamathee *et al*., 2023). The classification of F1 score measures takes precision and recall into account. It is calculated as the harmonic mean of precision and recall. To understand the performance of the CNN model in classification, all the common metrics: Precision, recall, and F1-score are calculated using the following Eqs. (35):

$$Precision = \frac{\sum TPi}{\sum TPi + \sum FPi} \tag{3}$$

$$Recall = \frac{\sum TPi}{\sum TPi + \sum FNi} \tag{4}$$

$$F1 = \frac{2 \times \Pr ecision \times \operatorname{Re} call}{\Pr ecision + \operatorname{Re} call} \tag{5}$$

Results and Discussion

Comparative Experiments

To compare the overall performance of MobileNet, the proposed method is compared with 2 deep learning and 4 traditional machine learning, including VGG16, Resnet101, KNN, SVM, decision tree, and random forest. Precision, recall, and F1 are used as the evaluation metrics. It is evident that MobileNet has a high precision (0.942), recall (0.947), and F1 (0.942), better than 3 state-of-the-art deep learning and 4 traditional machine learning Table (2). Before and after optimization, the precision, recall, and F1-score of the improved MobileNet architecture improved by 6.40, 3.25 and 5.99% in respectively Fig. (4). The new models are fine-tuned, resulting in significant enhancements in training precision, generalization ability, and robustness.

In addition, MobileNet has a smaller model size, fewer parameters, and a smaller trained model size Table (3).

Table 2: Quantitative evaluation metrics with different deep learning and traditional machine learning

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
KNN	24.56	24.7	24.7	23.9
SVM	25.56	25.7	26.4	25.1
Decision tree	24.63	24.7	24.6	24.5
Random forest	26.76	26.9	26.7	26.4
VGG16	64.58	62.5	72.4	57.9
Resnet101	91.97	92.0	94.4	87.5
Enhanced MobileNet	95.00	94.2	94.7	94.2

Table 3: The model size of the main model

Model	Model size (MB)	Parameters (million)
VGG16	57.157	149.836
Resnet101	166.800	437.258
MobileNet	14.358	37.640

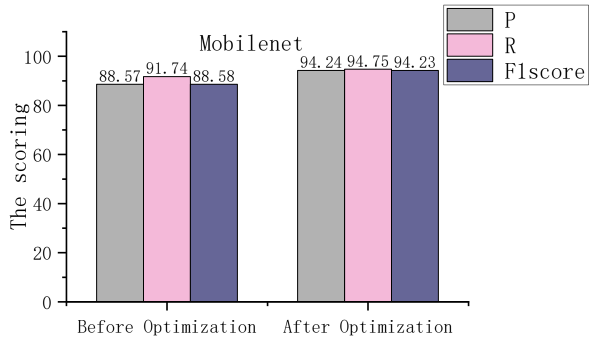


Fig. 4: Capsule classification

Selection of K Fold for MobileNet Model

The confusion matrix depicting all K = 10 fold results is shown in Fig. (5). The analysis of the confusion matrix reveals the advantages of values along the principal diagonal, representing true positives and true negatives, while off-diagonal values indicate instances of false positives and false negatives. An optimal classification is denoted by a metric. In this study, the achieved values approximate 0.9, closely approaching the ideal value of 1. This indicates that the proposed method has effectively recognized the majority of images within each class, underscoring its robust classification performance.

Effect of Learning Rate and Batch Size on Detecting Results

Learning rate and batch size are hyperparameters in cross-validation that significantly impact the effectiveness of a neural network on a dataset. Both excessively high and low learning rates can decrease the accuracy of recognition. The learning rate determines the magnitude of weight adjustments at each step of gradient descent, directly influencing the speed of model convergence. A higher learning rate can accelerate the model convergence initially but may cause oscillations that hinder the optimal solution. Conversely, a lower learning rate can reduce convergence. To determine the optimal learning rate, this study selected a smaller batch size and a specific number of neurons in fully connected layers (batch size = 64, neurons = 512). Small batches can slow down training, as they help update model parameters more frequently. The choice of neuron count depends on dataset complexity and task requirements. Too few neurons will lead to underfitting, while too many will cause overfitting. Table (4) shows different tested learning rates during training to balance between learning rate and detection accuracy.

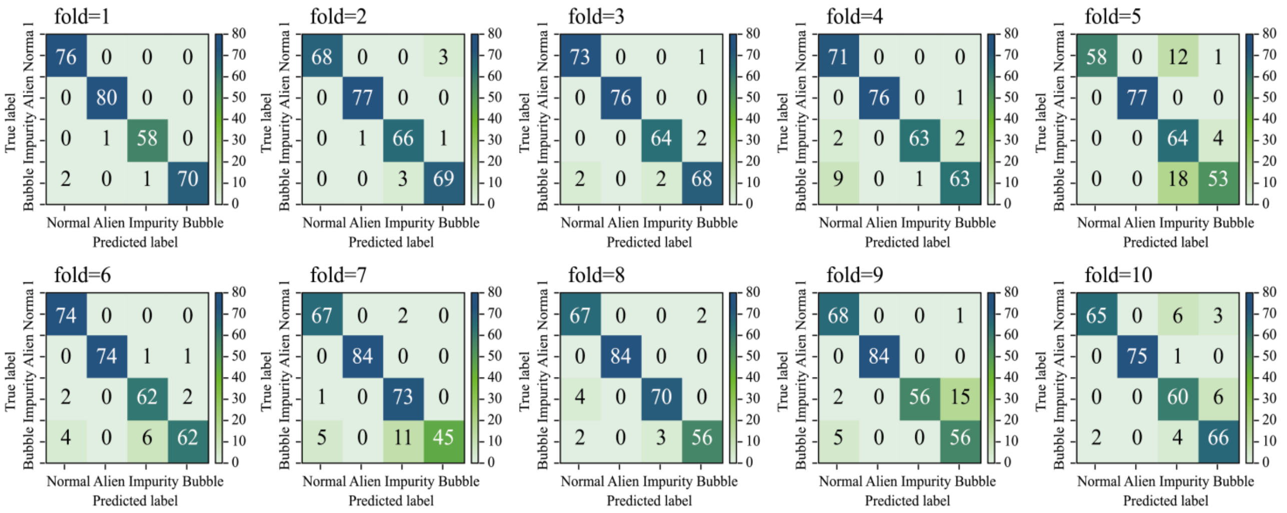


Fig. 5: Confusion matrix of MobileNet model in the testing data for all the 10 folds

Table 4: Impact of learning rate on recognition performance

Learning_rate	Train_accuracy (%)	Val_accuracy (%)	Test_accuracy (%)	Train_loss	Val_loss
0.01	96.03	93.72	95.00	0.21	0.27
0.001	95.97	92.95	92.29	0.21	0.36
0.0001	95.79	93.18	94.57	0.21	0.30

Table 5: Impact of batch size on classification outcomes

Bath _size	Trian_accuracy (%)	Val_accuracy (%)	Test_accuracy (%)	Train_loss	Val_accuracy
32	94.68	93.86	91.65	0.23	0.36
64	96.03	93.72	95.00	0.21	0.27
96	96.28	93.25	93.23	0.22	0.38
128	96.26	92.04	91.40	0.24	0.57

As the learning rate decreases, the detection accuracy steadily improves, although excessively small rates can lead to declines in accuracy. Clearly, a learning rate of around 0.01 achieves the highest detection accuracy, indicating the model reaches its optimal state at this point.

Batch size is the amount of sample data fed into the network at once when training the model. In classification tasks, selecting an appropriate batch size is crucial for improving the network model's training speed and accuracy. When the batch size is excessively small, it may result in unstable convergence of the network and a slower convergence rate, causing fluctuations in the loss value. Alternatively, an extremely large batch size results in excessive memory consumption and prolonged computations, leading to local optimum (Nan *et al*., 2022).

This study is conducted under a learning rate of 0.01 and 512 neurons. The batch sizes of 32, 64, 96, and 128 are compared all of which are common choices in deep learning (Krizhevsky *et al*., 2017). According to the results in Table (5), the loss function reaches its minimum, and setting the batch size to 64 significantly improves the accuracy of the test set to its highest level.

Figures (6-7) illustrate the impact of the learning rate on the recognition outcomes. Figures (6a-b) depict the accuracy of the training and validation datasets across varying learning rates, respectively. Notably, at a learning rate of 0.01, the accuracy rate surpasses that of another learning rate with the same number of iterations. In addition, the accuracy gradually improves as the training iterations increase, reaching a slow and stable growth point of around 30 iterations. Upon achieving approximately 70 iterations, the recognition accuracy exhibits gradual improvement and subsequently stabilizes. Figures (7a-b) present the loss curves for the training and validation datasets across varying learning rates, respectively. Clearly, after 10 iterations, a minimal loss is observed in the networks when employing a learning rate of 0.01. As additional iterations are conducted, the loss value consistently diminishes, tending towards zero.

Figures (8-9) illustrate the impact of batch size on recognition outcomes. Specifically, Fig. (8a-b) displays the accuracy trends for both the training and validation datasets with varying batch sizes. Particularly, classification correctness improves progressively and stabilizes after approximately 30 iterations. Similarly, Fig. (9a-b) illustrates the training and validation loss curves for varying batch sizes. Figure (9) indicates that larger batch sizes correspond to higher initial loss values. Conversely, smaller batch sizes result in erratic accuracy curves, signifying network instability.

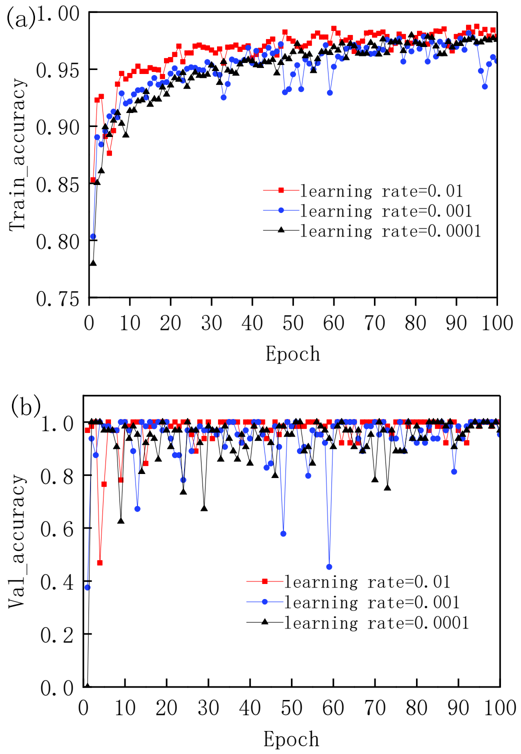


Fig. 6: Curve depicting recognition accuracy across varying learning rates; (a) train_accuracy; (b) validation_accuracy

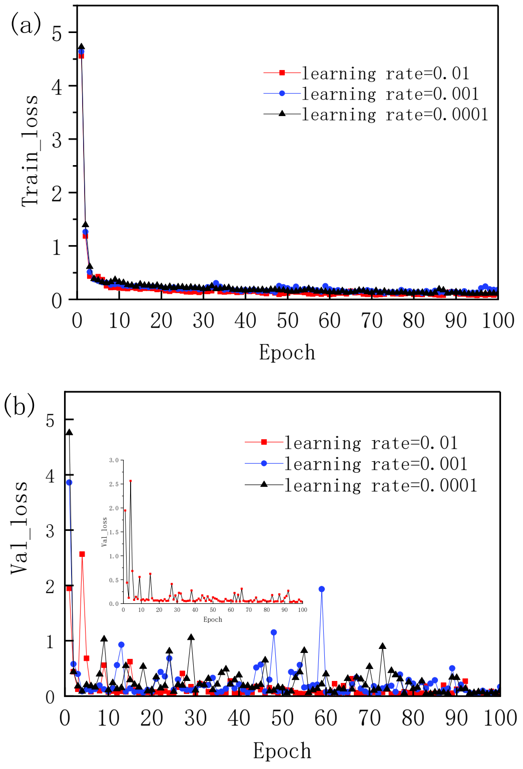


Fig. 7: Loss trajectory with varying learning rates; (a) train_loss; (b) validation_loss

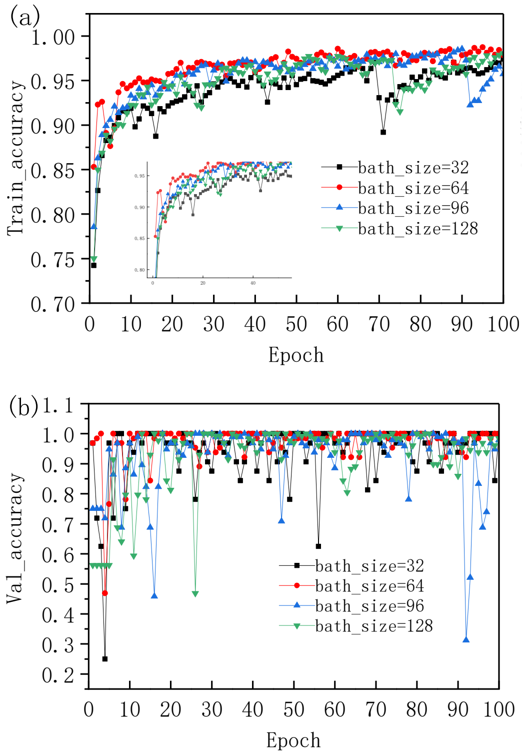


Fig. 8: Accuracy curve with different batch_size; (a) train_accuracy; (b) validation _accuracy

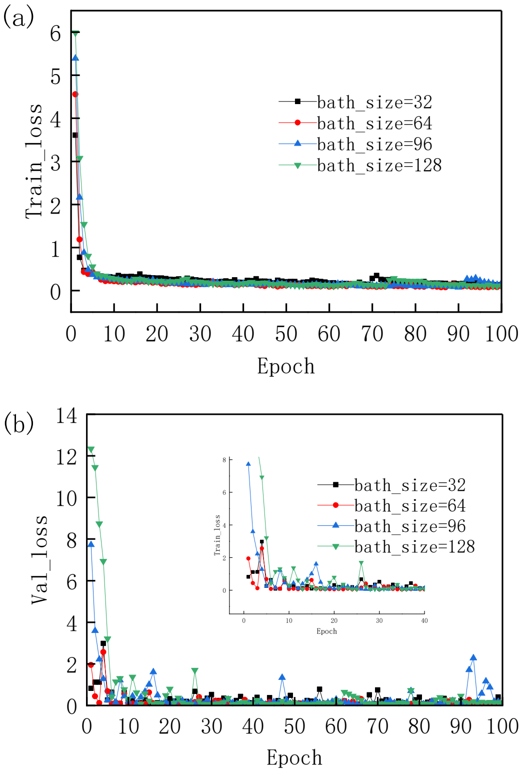


Fig. 9: Loss curve with different batch_size; (a) train_loss; (b) validation _loss

Effect of MobileNet Structure on Detecting Results

Figures (10-11) show the effect of neurons in the fully connected layer on recognition results. Figures (10a-b) show the accuracy curves of the training set and validation set, respectively. It is observed that the number of neurons reaches 256 and 512 and the curve converges well. Figures (11a-b) depict the loss trajectories for the training and validation datasets, respectively. As the training iterations increase, the accuracy at 256 layers and 512 layers surpasses that at 1024 Fig. (11a). However, the validation accuracy at 256 layers becomes unstable in the later stages of the training, while at 512 layers, it remains consistently stable throughout the training period Table (6). This shows that the 512 layer converges well during the training process.

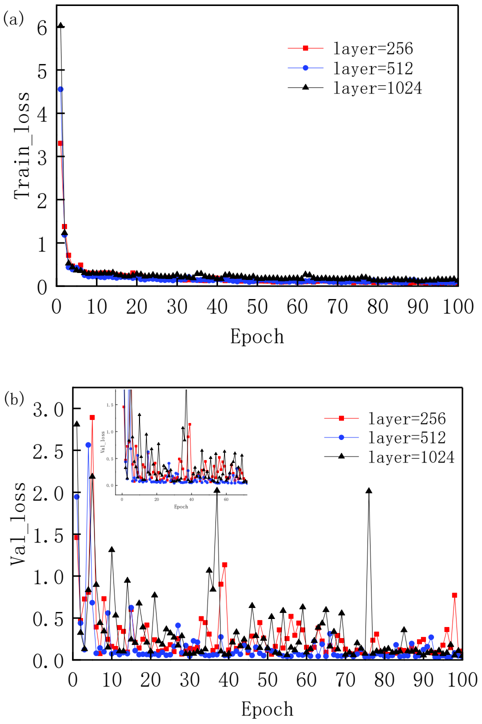
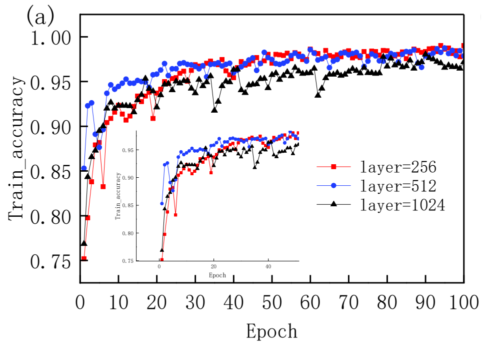


Fig. 10: Loss curve under different layers (256, 512, 1024); (a) train_loss; (b) validation _loss



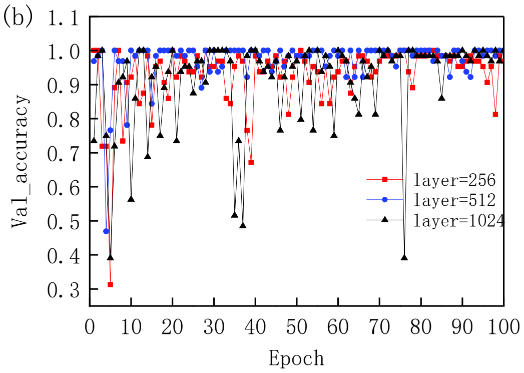


Fig. 11: Accuracy curve under different layers (256, 512, 1024); (a) Train_accuracy; (b) Validation _accuracy

Table 6: Impact of neuron count in the fully connected layer on recognition performance

Number of neurons	Train_accuracy (%)	Val_accuracy (%)	Test_accuracy (%)	Train_loss	Test_loss
256	95.98	93.15	93.95	0.19	0.28
512	96.03	93.72	95.00	0.21	0.27
1024	95.62	92.86	94.05	0.36	0.24

Conclusion

Anew approach is proposed for detecting Capsule Defects by using transfer learning models such as MobileNet, VGG16, and Resnet101. The primary conclusions can be summarized as follows:

1. MobileNet has fewer parameters, runs faster, lowers training costs, and is well-suited for capsule defect detection in medical applications
2. The parameters and structure have an effect on the performance of MobileNet. The optimal settings, including a K value of 10, a batch size of 64, a learning rate of 0.001, and a fully connected layer size of 512, can yield the best performance
3. The proposed MobileNet architecture has shown significant effectiveness, resulting in a 7.9% improvement in accuracy over the optimized model, and yields precision, recall, and F1 scores of 94.24, 94.75, and 94.23%, respectively. It is crucial to study the impact of hyperparameters on prediction accuracy through cross-validation. Additionally, the fully connected layer situated at the top of the depth model also has a significant impact on prediction accuracy. It has been proven that using transfer learning can effectively prevent overfitting and improve accuracy. In this study, we constructed a new top-level architecture based on MobileNet and achieved significant accuracy improvement through fine-tuning. The proposed model has demonstrated high classification

accuracy on specific capsule datasets. However, due to the noise in other datasets, the performance may decrease. Besides the dataset's diversity, the imbalanced distribution across various categories can adversely affect the training process, leading to reduced accuracy. Another challenge is the time-consuming task of selecting suitable hyperparameters, including factors like batch size and learning rate, to enhance capsule classification accuracy through cross-validation. Due to differences in the dataset, the optimal hyperparameter selection may vary and it is necessary to re-evaluate these parameters when retraining the dataset

In the future, the potential for optimizing fixed layer depths in transfer learning is enormous. This approach utilizes pre-trained models and prior knowledge to significantly accelerate the learning process of new tasks, especially suitable for scenarios with limited data or constrained resources. By adjusting the layer depth based on task complexity and data characteristics, the model performance has been effectively improved, solving the problems of data imbalance and sample scarcity, thereby reducing costs and improving generalization ability.

Acknowledgment

The author expresses gratitude to Tianjin Tasly Group Co., Ltd. for their support and thanks Professor Huang for their guidance.

Funding Information

The authors have no support or funding to report.

Author’s Contributions

Wenqing Bian: Designed and performed the experiments, and work.

Yongzhi Huang: Designed the experiments and revised the manuscript.

Baolian Shan and Yanting Chang: Participated in collecting the materials related to the experiment.

Ethics

The authors declare their responsibility for any ethical issues that may arise after the publication of this manuscript.

Conflict of Interest

The authors declare that they have no competing interests. The corresponding author affirms that all of the authors have read and approved the manuscript.

References

Akil, M., Saouli, R., & Kachouri, R. (2020). Fully automatic brain tumor segmentation with deep learning-based selective attention using overlapping patches and Multi-Class Weighted Cross-Entropy. *Medical Image Analysis*, *63*, 101692. https://doi.org/10.1016/j.media.2020.101692

Chang, H. C., Yu, L. W., Liu, B. Y., & Chang, P. C. (2024). Classification of the implant-ridge relationship utilizing the mobilenet architecture. *Journal of Dental Sciences*, *19*(1), 411–418. https://doi.org/10.1016/j.jds.2023.08.002

Chawla, T., Mittal, S., & Azad, H. K. (2024). MobileNet-GRU fusion for optimizing diagnosis of yellow vein mosaic virus. *Ecological Informatics*, *81*, 102548. https://doi.org/10.1016/j.ecoinf.2024.102548

Ehtisham, R., Qayyum, W., Camp, C. V., Plevris, V., Mir, J., Khan, Q. U. Z., & Ahmad, A. (2024). Computing the characteristics of defects in wooden structures using image processing and CNN. *Automation in Construction*, *158*, 105211. https://doi.org/10.1016/j.autcon.2023.105211

Heidari, M., Mirniaharikandehei, S., Khuzani, A. Z., Danala, G., Qiu, Y., & Zheng, B. (2020). Improving the performance of CNN to predict the likelihood of COVID-19 using chest x-ray images with Preprocessing Algorithms. *International Journal of Medical Informatics*, *144*, 104284. https://doi.org/10.1016/j.ijmedinf.2020.104284

Hou, H., Wang, X., Zhang, D., & Zhang, H. (2011). An effective method for extracting capsule by color Image Processing. *2011 International Conference of Information Technology, Computer Engineering and Management Sciences, Nanjing, China, IEEE*, 295–298. https://doi.org/10.1109/icm.2011.111

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*. https://doi.org/10.48550/arXiv.1704.04861

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Mohapatra, S., Abhishek, N. V. S., Bardhan, D., Ghosh, A. A., & Mohanty, S. (2021). Comparison of MobileNet and Reset CNN Architectures in the CNN-Based Skin Cancer Classifier Model. In Mohanty, S.N., Nalinipriya, G., Jena, O.P. and Sarkar, A. (Eds.), *Machine Learning for Healthcare Applications*, John Wiley and Sons, 169-186. https://doi.org/10.1002/9781119792611.ch11

Nan, Y., Ju, J., Hua, Q., Zhang, H., & Wang, B. (2022). A-MobileNet: An approach of facial expression recognition. *Alexandria Engineering Journal*, *61*(6), 4435–4444. https://doi.org/10.1016/j.aej.2021.09.066

Naranjo-Torres, J., Mora, M., Hernández-García, R., Barrientos, R. J., Fredes, C., & Valenzuela, A. (2020). A Review of Convolutional Neural Network Applied to Fruit Image Processing. *Applied Sciences*, *10*(10), 3443. https://doi.org/10.3390/app10103443

Nguyen, T. P., Choi, S., Park, S. J., Park, S. H., & Yoon, J. (2021). Inspecting Method for Defective Casting Products with Convolutional Neural Network (CNN). *International Journal of Precision Engineering and Manufacturing-Green Technology*, *8*(2), 583–594. https://doi.org/10.1007/s40684-020-00197-4

Pławiak, P., & Acharya, U. R. (2020). RETRACTED ARTICLE: Novel deep genetic ensemble of classifiers for arrhythmia detection using ecg signals. *Neural Computing and Applications*, *32*(15), 11137–11161. https://doi.org/10.1007/s00521-018-03980-2

Rahman, M. H., Jannat, M. K. A., Islam, M. S., Grossi, G., Bursic, S., & Aktaruzzaman, M. (2023). Real-Time Face Mask Position Recognition System Based on MobileNet Model. *Smart Health*, *28*, 100382. https://doi.org/10.1016/j.smhl.2023.100382

Surono, S., Afitian, M. Y. F., Setyawan, A., Arofah, D. K. E., & Thobirin, A. (2023a). Comparison of CNN Classification Model Using Machine Learning with Bayesian Optimizer. *High-tech and Innovation Journal*, *4*(3), 531–542. https://doi.org/10.28991/hij-2023-04-03-05

Surono, S., Rivaldi, M., Dewi, D. A., & Irsalinda, N. (2023b). New Approach to Image Segmentation: U-Net Convolutional Network for Multiresolution CT Image Lung Segmentation. *Emerging Science Journal*, *7*(2), 498–506. https://doi.org/10.28991/esj-2023-07-02-014

Sutaji, D., & Yıldız, O. (2022). LEMOXINET: Lite Ensemble MobileNet V2 and Xception Models to Predict Plant Disease. *Ecological Informatics*, *70*, 101698. https://doi.org/10.1016/j.ecoinf.2022.101698

Thanathamathee, P., Sawangarreerak, S., Kongkla, P., & Nizam, D. N. M. (2023). An Optimized Machine Learning and Deep Learning Framework for Facial and Masked Facial Recognition. *Emerging Science Journal*, *7*(4), 1173–1187. https://doi.org/10.28991/esj-2023-07-04-010

Zhou, J., He, J., Li, G., & Liu, Y. (2020). Identifying Capsule Defect Based on an Improved Convolutional Neural Network. *Shock and Vibration*, *2020*(1), 8887723. https://doi.org/10.1155/2020/8887723

Zhou, W., Wang, H., & Wan, Z. (2022). Ore Image Classification Based on Improved CNN. *Computers and Electrical Engineering*, *99*, 107819. https://doi.org/10.1016/j.compeleceng.2022.107819

Zhu, Z.T., Z., Yu, X.Y., Huang, L.Q.,Wu, D. (2011). Fast Capsule Image Segmentation Based on Linear Region Growing. *2011 IEEE International Conference on Computer Science and Automation Engineering, Shanghai, China, IEEE*, 99–103. https://doi.org/10.1109/csae.2011.5952433