

Research Article

Medical Image Classification Algorithm Based on Weight Initialization-Sliding Window Fusion Convolutional Neural Network

Feng-Ping An ^{1,2}

¹School of Physics and Electronic Electrical Engineering, Huaiyin Normal University, Huaian 223300, China

²School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China

Correspondence should be addressed to Feng-Ping An; anfengping@163.com

Received 4 July 2019; Accepted 17 September 2019; Published 20 October 2019

Academic Editor: Yong Xu

Copyright © 2019 Feng-Ping An. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to the complexity of medical images, traditional medical image classification methods have been unable to meet actual application needs. In recent years, the rapid development of deep learning theory has provided a technical approach for solving medical image classification tasks. However, deep learning has the following problems in medical image classification. First, it is impossible to construct a deep learning model hierarchy for medical image properties; second, the network initialization weights of deep learning models are not well optimized. Therefore, this paper starts from the perspective of network optimization and improves the nonlinear modeling ability of the network through optimization methods. A new network weight initialization method is proposed, which alleviates the problem that existing deep learning model initialization is limited by the type of the nonlinear unit adopted and increases the potential of the neural network to handle different visual tasks. Moreover, through an in-depth study of the multicolumn convolutional neural network framework, this paper finds that the number of features and the convolution kernel size at different levels of the convolutional neural network are different. In contrast, the proposed method can construct different convolutional neural network models that adapt better to the characteristics of the medical images of interest and thus can better train the resulting heterogeneous multicolumn convolutional neural networks. Finally, using the adaptive sliding window fusion mechanism proposed in this paper, both methods jointly complete the classification task of medical images. Based on the above ideas, this paper proposes a medical classification algorithm based on a weight initialization/sliding window fusion for multilevel convolutional neural networks. The methods proposed in this study were applied to breast mass, brain tumor tissue, and medical image database classification experiments. The results show that the proposed method not only achieves a higher average accuracy than that of traditional machine learning and other deep learning methods but also is more stable and more robust.

1. Introduction

With the rapid development of computer and medical imaging, medical imaging techniques, such as computed tomography (CT) and magnetic resonance imaging (MRI), can noninvasively reflect the physiological state of tissues and organs in the human body. These techniques have gradually become indispensable tools in medical research, clinical diagnosis, and surgical planning [1–3]. While these new technologies have advanced medical theory and practice, they have also raised new

issues; for example, doctors now need to classify diagnostic results. Automatic image classification techniques can understand the image content to a certain extent; for example, they can effectively identify lesion areas in medical images and assist doctors in carrying out efficient diagnoses [3, 4]. However, many types of medical images exist, and distinguishing the categorical information in these medical images often requires different processing and analysis methods [5–7]. At present, medical image classification is mostly based on pattern recognition methods, in which a classification

model is trained to identify and distinguish medical images. Medical image classification can generally be divided into supervised classification methods and unsupervised classification methods [8–10].

Supervised classification means that the processed samples are labeled in advance; the classification model is trained using the labeled image features, and the classification categories are usually prespecified. Supervised classification methods mainly include the K-nearest neighbor algorithm [11], Bayesian models [12], logistic regression [13], neural networks [14], and support vector machines [15]. Among these methods, medical image classification methods based on neural networks tend to perform the best [16–19]. Abbass [20] proposed a neural network based on the Pareto differential evolution algorithm to classify breast cancer on the Wisconsin Breast Cancer Database (WBCD) that obtained a better classification performance than those of traditional neural networks. Karabatak and Ince [21] used association rules for dimension reduction, reducing nine features to four, and then classified them using an artificial neural network (ANN). The accuracy of 2-fold cross-validation on breast cancer in the WBCD reached 90%. However, this kind of method cannot adaptively match the characteristic information in the medical images themselves, which leads to large classification effect differences on different medical images.

Unsupervised classification methods automatically distinguish different categories based on the similarities between samples without requiring pre-labeled samples. Unsupervised learning is essentially a clustering process. The typical unsupervised methods include K-means clustering [22], fuzzy C-means clustering [23], and principal component analysis (PCA) [24]. To solve the problem that lesion contours cannot be accurately found in MRI images, in 2010, Juan proposed a color conversion segmentation algorithm with a K-means clustering technique that added a color-based segmentation operation [22]. This method achieved a higher accuracy but with a poorer segmentation effect than the fuzzy segmentation algorithm for MRI data proposed by Zhang and Chen in 2004 [23]. In 2012, Singh and Kaur proposed a PCA-based method to automatically classify MRI and natural images. Their experiments showed that the classification accuracy reached 91% [24]. Although this method is relatively simple to implement, it has achieved certain effects in image recognition and classification applications. However, the classification effect of this kind of method has great differences when classifying and recognizing different medical images, and it cannot adaptively classify and recognize medical images based on their characteristics.

In recent years, medical image classification research based on deep learning has once again attracted scholarly attention. Deep learning [25] is an organic combination of supervised and unsupervised methods. It primarily relies on unsupervised learning to train deep neural networks and then fine tunes them through supervised learning methods [26–29]. In the early stages of deep neural network

development, medical image recognition focused on unsupervised pretraining methods, such as stacked autoencoders (SAEs) and restricted Boltzmann machines (RBMs). For example, in 2013, Brosch and Tam [30] used a deep belief network to classify neuroimages. Plis et al. [31] used a deep belief network and an SAE to determine whether brain-based MRI can be used to diagnose whether a patient has Alzheimer's disease. Cheng et al. [32] used an SAE for the intelligent identification of breast ultrasound lesions; the performance of this method was significantly improved, compared with those of conventional methods. Kallenberg et al. [33] used convolutional SAEs to extract features from unlabeled breast cancer X-ray images. The main difference between convolutional SAEs and convolutional neural networks (CNNs) is the use of SAEs for pretraining [34–37]. For such tasks, it is often necessary to combine local information about the appearance of the lesion with global context information about the location of the lesion to determine a more precise classification [37, 38]. Such requirements are difficult to achieve through common deep learning architectures. Other researchers have tried to solve this problem using multinet network branching architecture. For example, in 2015, Shen et al. [39] constructed three CNNs and connected their outputs to form the final feature vector. In 2016, Kawahara and Hamarneh [40] attempted to classify skin lesions using multibranch CNNs, each of which worked on a different image resolution. In 2017, Esteva et al. [7] used 129450 clinical image datasets to train CNN models. The results showed that the classification level of skin cancer by the CNNs reached those of dermatologists. In 2018, Bidart et al. [41] designed a method to perform automatic localization of breast cancer tissue sections using fully convolutional neural networks (FCNNs). The breast cancer tissue sections were divided into three types: lymphocytes, benign epithelial cells, and malignant epithelial cells, and the classification accuracy rate reached 94.6%. In summary, deep learning models have been widely applied to various medical image classification tasks. However, medical images are not the same as natural images. Thus, constructing a CNN that achieves a better performance than other intelligent classification methods while considering the specific characteristics of medical images is a difficult problem. Moreover, because the network depth affects the nonlinear modeling ability of the CNN, improper initialization will cause the deep network to have difficulty converging, and no good initialization method exists. To address this issue, this paper first improves the nonlinear modeling ability of the network from the network optimization perspective; then, it improves the generalizability of the solution after the convergence domain has been improved. Furthermore, a new network weight initialization method is proposed. This method alleviates the problem that the initialization theory of existing deep learning is limited by the type of nonlinear unit; it provides more choice for the structural design of deep networks and increases the potential of neural networks to handle different visual tasks. In addition, to make full use of the characteristics of medical images, this paper studies the

framework of multicolumn CNNs and constructs CNN models with different structures that can better adapt to the characteristics of the medical images of interest. Then, different CNN models are trained on the same dataset. Finally, the trained heterogeneous, multicolumn CNN is combined with an adaptive sliding window fusion mechanism proposed in this paper; this combination completes the medical image classification task. Based on the above ideas, this paper proposes a medical classification algorithm based on weight initialization/sliding window fusion for multilevel CNNs.

Section 2 of this paper mainly presents the weight initialization algorithm based on the deep learning model proposed in this paper. Section 3 systematically describes the sliding window fusion CNN model proposed in this paper. Section 4 introduces a medical classification algorithm based on weight initialization and multilevel CNN sliding window fusion. Section 5 analyzes the proposed medical image classification algorithm and compares it with mainstream medical image classification algorithms. Finally, Section 6 summarizes and discusses the full text.

2. Deep Learning Model Weight Initialization Method-Based Adaptive Taylor

The problem of classifying medical images is essentially one of the identifying features after their extraction. Although shallow networks can implement complex nonlinear transformations that convert input to output in a system, high computational complexity is required to achieve nonlinear representation capabilities similar to those of deep networks. Therefore, under the condition of corresponding increase in computational complexity, increasing network depth is the most reliable method to solve medical image classification problems. However, optimizing deep networks is a difficult problem in deep learning, especially given the exploding gradient problem caused by improper initialization. Of the existing methods, the Microsoft Research Asia (MSRA) method has good convergence and generalizability, but its disadvantage is that it is limited to a specific network type. In view of this shortcoming, this section proposes a new initialization method that aims to improve the convergence and generalization ability of the model through optimization techniques. The initialization method proposed in this paper combines the advantages of the above two methods to avoid their shortcomings, and it is essentially an extension of the MSRA method. Therefore, this section is divided into two parts: an introduction to the MSRA method and a description of the proposed method.

2.1. MSRA Method. For the i th convolutional layer, an output pixel y_i can be expressed as in [42]:

$$y_i = w_i \times x_i + b_i, \quad (1)$$

where y_i is a random variable and w_i and x_i are mutually independent random vectors (vectors are shown in bold). The offset term b_i is initialized with zeros. b_i is expressed as the offset term of the i th convolutional layer. The following formula gives the relationship between the variance of y_{i-1} and the variance of y_i :

$$\text{Var}(y_i) = \text{Var}(w_i \times x_i + b_i), \quad (2)$$

$$\text{Var}(y_i) = \text{Var}(w_i x_i), \quad (3)$$

$$\text{Var}(y_i) = k_i^2 c_i \text{Var}(w_i x_i), \quad (4)$$

where k_i represents the convolution kernel size, c_i is the number of channels input by the convolutional layer, and w_i and x_i are independent random variables. Formula (4) is established when the elements in the random vector w_i in formula (3) and the elements in x_i are independently and identically distributed. When w_i is initialized with a zero-mean symmetric distribution, formula (4) is transformed as follows:

$$\text{Var}(y_i) = k_i^2 c_i \text{Var}(w_i) E(x_i^2). \quad (5)$$

The following shows the relationship between $\text{Var}(y_i)$ and $\text{Var}(y_{i-1})$. The key is that there is a nonlinear unit f between x_i and y_{i-1} :

$$x_i = f(y_{i-1}). \quad (6)$$

For sigmoid nonlinearity, formula (6) can be adjusted to

$$x_i = \frac{1}{1 + e^{-y_{i-1}}}. \quad (7)$$

Based on formula (7), the methods from [43–46] assume a linear relationship between x_i and y_{i-1} near the origin. This formula derives the Xavier initialization method. For rectified linear unit (ReLU) nonlinearity, formula (6) can be adjusted to

$$x_i = \max(0, y_{i-1}). \quad (8)$$

He et al. [42] extended Xavier initialization to the ReLU network using formula (8).

2.2. Adaptive Taylor Initialization Method. Substituting formula (7) into formula (5) yields the Xavier initialization method applicable to a sigmoid network. Similarly, substituting formula (8) into formula (5) yields the MSRA initialization method applicable to a ReLU network. Therefore, the analytical solution of the initialization is a function of the relationship between x_i and y_{i-1} . However, if the expansion of formula (6) has a higher-order term, formula (5) will be difficult to solve. Therefore, the current mainstream initialization method is theoretically not applicable to networks other than those using ReLU or sigmoid activation functions. In response to this problem, this paper introduces the Taylor formula and proposes a more general initialization method.

According to the definition of nonlinear elements by Gulcehre [47], a nonlinear element is a mapping from real space to real space and can be guided almost everywhere. That is, $f: R \rightarrow R$. For the convenience of derivation, we define a nonlinear unit using the following function:

$$x = f(y), \quad (9)$$

where x is the output of the nonlinear unit, y is the input of the nonlinear unit, and f exists almost everywhere in $y \in R$. Suppose $y \in R$ has a point ε such that $f(\varepsilon) = 0$. The n th ($n \geq 1$) order Taylor expansion for f at $y = \varepsilon$ is

$$x = f(\varepsilon) + f'(\varepsilon)(y - \varepsilon) + \frac{f''(\varepsilon)}{2!}(y - \varepsilon)^2 + \dots + \frac{f^n(\varepsilon)}{n!}(y - \varepsilon)^n + R_n(y), \quad (10)$$

which can be found by formula (5). However, when the order of formula (10) is 2, the above formula is difficult to solve. To solve this problem, this section uses the Taylor formula to approximate the nonlinear element by taking the linear term of $n \leq 1$ in formula (10). By definition, if f is continuous at $y = \varepsilon$ and can be guided, formula (10) can be simplified to

$$x = f(\varepsilon) + f'(\varepsilon)(y - \varepsilon) + R_n(y), \quad (11)$$

$$x \approx f(\varepsilon) + f'(\varepsilon)(y - \varepsilon). \quad (12)$$

If f is nondifferentiable at $y = \varepsilon$, but its lower gradient exists (for example, ReLU), then

$$x \approx \begin{cases} f(\varepsilon) + f'(\varepsilon_+)(y - \varepsilon), & y > \varepsilon, \\ f(\varepsilon) + f'(\varepsilon_-)(y - \varepsilon), & y \leq \varepsilon. \end{cases} \quad (13)$$

Formula (12) can be regarded as a special case of formula (13). Therefore, this paper is based on formula (12).

For any $y \in R$, if $f'(\varepsilon_+) = f'(\varepsilon_-) = 0$, then $x = f(\varepsilon)$. This property shows that the output of the nonlinear element is constant, which causes the CNN to lose discriminative power and thus needs to be ruled out.

When neither $f'(\varepsilon_+)$ or $f'(\varepsilon_-)$ are zero, then if the convolutional layer parameter w is initialized with a zero-mean symmetric distribution, b is initialized with a constant $(\varepsilon + f'(\varepsilon_+))/f'(\varepsilon_+)$ for $y \leq \varepsilon$. b is the offset term of a convolutional layer. For $y \leq \varepsilon$, to initialize $(\varepsilon + f'(\varepsilon_+))/f'(\varepsilon_+)$ with a constant, then $f(\varepsilon) + f'(\varepsilon_+)(y - \varepsilon)$ must have a symmetric distribution with a mean of ε . Let x have a probability density function $p(x)$. For convenience, recall that $y - \varepsilon$ is \hat{y} . Then,

$$\begin{aligned} E(x^2) &= \int_{-\infty}^{+\infty} x^2 p(x) dx \approx \int_{-\infty}^{\varepsilon} (f(\varepsilon) + f'(\varepsilon_-)\hat{y})^2 \\ &\quad + p(f(\varepsilon) + f'(\varepsilon_-)\hat{y})d(f(\varepsilon) + f'(\varepsilon_-)\hat{y}) \\ &\quad + \int_{\varepsilon}^{+\infty} (f(\varepsilon) + f'(\varepsilon_+)\hat{y})^2 + p(f(\varepsilon) + f'(\varepsilon_+)\hat{y})d \\ &\quad (f(\varepsilon) + f'(\varepsilon_+)\hat{y}) \\ &= \frac{1}{2}E[(f(\varepsilon) + f'(\varepsilon_-)\hat{y})^2] + \frac{1}{2}E[(f(\varepsilon) + f'(\varepsilon_+)\hat{y})^2] \\ &= \frac{1}{2}E(f^2(\varepsilon) + 2f(\varepsilon)f'(\varepsilon_-)\hat{y} + f'^2(\varepsilon_-)\hat{y}^2) \\ &\quad + \frac{1}{2}E(f^2(\varepsilon) + 2f(\varepsilon)f'(\varepsilon_+)\hat{y} + f'^2(\varepsilon_+)\hat{y}^2) \\ &= \frac{1}{2}(f^2(\varepsilon) + 2f(\varepsilon)f'(\varepsilon_-)E(\hat{y}) + f'^2(\varepsilon_-)E(\hat{y}^2)) \\ &\quad + \frac{1}{2}(f^2(\varepsilon) + 2f(\varepsilon)f'(\varepsilon_+)E(\hat{y}) + f'^2(\varepsilon_+)E(\hat{y}^2)) \\ &= \frac{1}{2}(f^2(\varepsilon) + 2\varepsilon f(\varepsilon) + f'^2(\varepsilon_-))(\text{Var}(\hat{y}^2) + E(\hat{y}^2)) \\ &\quad + \frac{1}{2}(f^2(\varepsilon) + 2\varepsilon f(\varepsilon) + f'^2(\varepsilon_+))(\text{Var}(\hat{y}^2) + E(\hat{y}^2)) \\ &= \frac{1}{2}(f^2(\varepsilon) + 2\varepsilon f(\varepsilon) + \varepsilon^2 + f'^2(\varepsilon_-))(\text{Var}(\hat{y}^2)) \\ &\quad + \frac{1}{2}(f^2(\varepsilon) + 2\varepsilon f(\varepsilon) + \varepsilon^2 + f'^2(\varepsilon_+))(\text{Var}(\hat{y}^2)) \\ &= f^2(\varepsilon) + 2\varepsilon f(\varepsilon) + \frac{1}{2}(f'^2(\varepsilon_-) + f'^2(\varepsilon_+))\text{Var}(\hat{y}^2). \end{aligned} \quad (14)$$

Because $\text{Var}(\hat{y}) = \text{Var}(y - \varepsilon) = \text{Var}(y)$, formula (14) can be simplified to

$$E(x^2) \approx f^2(\varepsilon) + 2\varepsilon f(\varepsilon) + \frac{1}{2}(f'^2(\varepsilon_-) + f'^2(\varepsilon_+))\text{Var}(y). \quad (15)$$

Let $P = f^2(\varepsilon) + 2\varepsilon f(\varepsilon)$. Then, formula (15) can be substituted into formula (5), obtaining

$$\begin{aligned} \text{Var}(y_i) &\approx k_i^2 c_i \text{Var}(w_i) \left(P + \frac{1}{2}(f'^2(\varepsilon_-) + f'^2(\varepsilon_+)) \right) \text{Var}(y_{i-1}) \\ &= P k_i^2 c_i \text{Var}(w_i) + \frac{1}{2}(f'^2(\varepsilon_-) + f'^2(\varepsilon_+)) k_i^2 c_i \text{Var}(w_i) \\ &\quad \text{Var}(y_{i-1}). \end{aligned} \quad (16)$$

The relationship between the output variance of the L th layer and the output variance of the 1st layer is as follows:

$$\text{Var}(y_L) \approx P \sum_{j=2}^L \left(N^{L-j} \prod_{i=j}^L M_i \right) + \text{Var}(y_1) \prod_{i=2}^L M_i N, \quad (17)$$

where $M_i = k_i^2 c_i \text{Var}(w_i)$ and $N = (1/2)(f'^2(\varepsilon_-) + f'^2(\varepsilon_+))$. Similarly, we can see that $f'(\varepsilon_-)$ and $f'(\varepsilon_+)$ are 0.

Formula (17) shows that the size of $\text{Var}(y_L)$ is related to the exponential increase in L . For the nonlinear unit $f: R \rightarrow R$, if there is no point ε , let $f(\varepsilon) = 0$. Then, when L is large, $\text{Var}(y_L)$ will also be large. This phenomenon causes the output of the softmax layer to overflow, and the network cannot converge. For most nonlinear elements, the function f is usually designed to pass through the origin, i.e.,

$$f(0) = 0. \quad (18)$$

Take $\varepsilon = 0$, that is, perform a Taylor expansion on f at $y = 0$. Here, the initial value of b becomes 0; then, formula (17) can be simplified to

$$\text{Var}(y_L) \approx \text{Var}(y_1) \prod_{i=2}^L \frac{1}{2} (f'^2(0_-) + f'^2(0_+)) k_i^2 c_i \text{Var}(w_i). \quad (19)$$

It can be seen that there is a scaling factor $\prod_{i=2}^L (1/2)(f'^2(0_-) + f'^2(0_+)) k_i^2 c_i \text{Var}(w_i)$ between the output variance of the L th layer and the output variance of the 1st layer. According to [42], a reasonable initialization method can avoid exponentially increasing or decreasing amplitude of the signal during forward transmission. The sufficient conditions to establish the above conclusions are as follows:

$$\frac{1}{2} (f'^2(0_-) + f'^2(0_+)) k_i^2 c_i \text{Var}(w_i) = 1. \quad (20)$$

Substituting formula (20) into formula (19) yields

$$\text{Var}(w_i) = \frac{2}{(f'^2(0_-) + f'^2(0_+)) k_i^2 c_i}, \quad (21)$$

which shows that if the variance in the convolution layer parameters satisfies the above relationship, then the amplitude of the input signal will not diverge or disappear during the forward transfer. This relationship further ensures the rationality of the gradient flow in the backward transfer process. Through the above analysis, the adaptive Taylor initialization method proposed in this paper is as follows:

$$\begin{cases} w_i \sim \left(0, \sqrt{\frac{2}{(f'^2(0_-) + f'^2(0_+)) k_i^2 c_i}} \right), \\ b = 0. \end{cases} \quad (22)$$

That is, the convolution parameter w is initialized with a Gaussian distribution $(0, \sqrt{2/((f'^2(0_-) + f'^2(0_+)) k_i^2 c_i)})$,

and the convolution layer offset b is initialized to 0. Except for $f(0) = 0$, this method does not define the specific functional form of f in the derivation process. Therefore, formula (21) is more relaxed than the MSRA and Xavier conditions; that is, it is suitable for a network with piecewise linearity, a network with piecewise exponential nonlinearity and even other types of networks. In addition, the existing initialization method can also be considered a special case of the method in this paper.

3. Sliding Window Fusion Method Based on Multilayer Convolutional Neural Network

3.1. Multilayer Convolutional Neural Network. A multilayer CNN is formed by multiple CNNs with different structures. A schematic diagram of a multilayer CNN is shown in Figure 1. This number of layers is a more optimized form obtained through the actual model training process in this paper. Generally, 20 layers or less will be selected because too many layers will affect the training time of the model.

During the convolution process, the convolutional layer affects the model as follows: first, the size of the convolution kernel determines the scale of the receptive field and affects the amount of feature information; second, the number of convolution kernels determines the richness of the feature information. In view of these properties, this section first constructs multiple CNNs with different convolution kernel sizes and numbers of feature maps and then trains each convolutional neural network, thereby achieving the purpose of fitting the training datasets to different network model structures. Finally, the outputs of the multilayer convolutional neural network are combined to form the final output, and better classification accuracy is achieved.

3.2. Sliding Window Fusion Method for Multilayer Convolutional Neural Networks. Different preprocessing methods were proposed for input images in [32] to obtain diverse network models. However, this paper is different from [32]; here, the diversity of multilayer CNNs is achieved by training multiple CNNs with different structures.

The classifier fusion methods mainly include the min, max, average, and product rules. These methods have a single function and many limitations. Therefore, this paper proposes a new classifier fusion method based on a sliding window. This method is a generalization of traditional classifier fusion methods, but it is more flexible and generalizable.

3.2.1. Sliding Window Fusion Process. In this paper, a 10-layer convolutional neural network is used as an example to describe the sliding window fusion method. The basic flow of the method is shown in Figure 2.

For an input medical image, after passing through each layer of the CNN during forward conduction, W_1 to W_{10} are the classification probabilities of each layer of the CNN for a certain category. The above probabilities are first sorted from low to high; then, a sliding window is applied to the sorted classification probability distribution to produce the

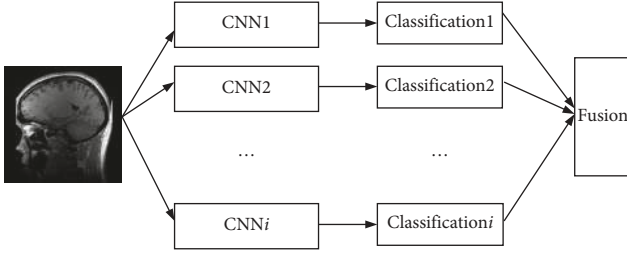


FIGURE 1: Structure of the multilayer convolutional neural network proposed in this paper.

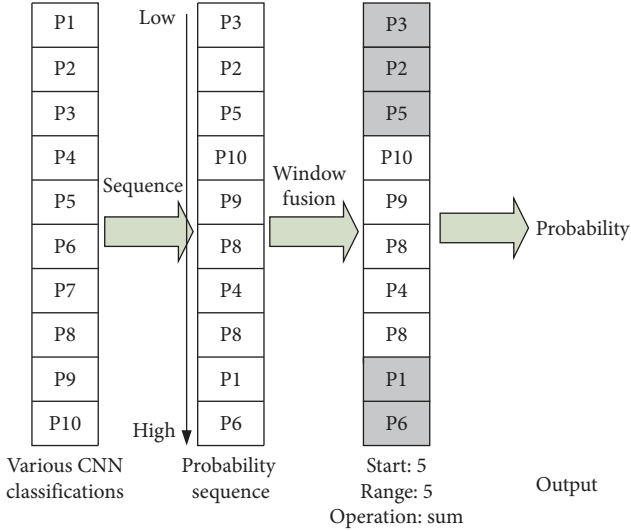


FIGURE 2: Schematic diagram of the sliding window fusion method for multilayer CNNs.

final classification result. This process mainly accomplishes the following tasks:

- (1) It determines which network classifications will be integrated into the subsequent results. This function is determined by the start and end ranges of the window (Range).
- (2) It determines the fusion of the selected classification results, which is determined by the parameter Operation. During the actual network fusion process, sliding window fusion is more flexible than the traditional single fusion method. The integration process is as follows:

(1) Parameter description

The input layer (P) $P(i, j)$ represents the prediction probability of the j th category of the i th column of the CNN, S is the starting position of the sliding window (Start), R is the range of the sliding window (Range), and O is the operation of the sliding window (operation, which can be set to “sum” or “product”). The output layer (T) $T(j)$ represents the prediction probability of the j th category after the sliding window is merged.

- (2) The number of columns in the convolutional neural network is $m \leftarrow \text{size}(P, 1)$.

- (3) The number of categories predicted is $n \leftarrow \text{size}(P, 2)$.
- (4) For category j to be cyclically processed from 1 to n , the predictions of the multicolumn CNN are spliced into a vector: $t \leftarrow [P(1, j), \dots, P(m, j)]$. The elements in vector t are sorted: $t_s \leftarrow \text{size}(P, 1)$.
- (5) When $o = \text{sum}$, $T(j) \leftarrow$ is accumulated from $t_s(h)$, and r elements are added. When $o = \text{product}$, $T(j) \leftarrow$, and r elements are multiplied from the beginning of $t_s(h)$.

Start, Range, and Operation are the three sliding window parameters. Start and Range determine which independent CNN classifications are used to calculate the final classification probability for the entire system. Start indicates the starting position of the sorted probabilities. Range determines how many categorical values are used for the calculation of the final classification. When Range exceeds the number of CNNs, the classification value continues to be selected from the first classified value after sorting. Sum and Product are the two methods that can be selected by the parameter Operation. Sum represents the summation of all the classification probabilities selected by the sliding window to attain the final classification ($M = \text{Start} + \text{Range} - 1$, $\mathcal{Y}_{\text{sort-CNN}_j}^i$ represents the sorted probability of sorting):

$$\mathcal{Y}_{\text{MCDNN}}^i = \sum_{j=\text{start}}^M \mathcal{Y}_{\text{sort-CNN}_j}^i \quad (23)$$

Product calculates the final classification probability by multiplying all the probabilities selected by the sliding window:

$$\mathcal{Y}_{\text{MCDNN}}^i = \prod_{j=\text{start}}^M \mathcal{Y}_{\text{sort-CNN}_j}^i \quad (24)$$

According to formulas (23) and (24), the method is a generalized classifier fusion method. For example, when the parameters of the sliding window start and range are both 1 $\mathcal{Y}_{\text{MCDNN}}^i = \mathcal{Y}_{\text{sort-CNN}_j}^i$, since $\mathcal{Y}_{\text{sort-CNN}_j}^i$ represents the probabilities sorted from low to high, and the sliding window fusion follows the traditional min rule case:

$$\mathcal{Y}_{\text{MCDNN}}^i = \min_{j \in [1, N]} \mathcal{Y}_{\text{CNN}_j}^i \quad (25)$$

After adjusting the sliding window to blend the parameters, the method is converted to other classifier methods, as shown in Table 1.

3.2.2. Parameter Acquisition Method. The parameters involved in the sliding window fusion method (Start, Range, and Operation) are obtained through exhaustion and training. Although the exhaustive method theoretically achieves the best classification effect, it has the following disadvantages: (1) the algorithm used in the exhaustive method to obtain the parameters is computationally expensive, with a complexity of $O(n^2)$, where n is the number of CNN models. (2) The exhaustive mode tests known data better than unknown data, which are less effective. Therefore, the exhaustive approach does not meet the actual testing requirements. In contrast, the training method first

TABLE 1: Special cases of the sliding window fusion method.

Start	Range	Operation	Fusion method
1	1	—	Min rule
N	1	—	Max rule
—	N	Sum	Average rule
—	N	Product	Product rule

obtains all the parameters on the training set and finally finds the optimal parameters for related testing. The complexity of this approach is still $O(n^2)$, but the complexity of the test is reduced to $O(1)$. The relevant parameters of the method are obtained by training, which is beneficial to improving the versatility and universality of the method. Therefore, the proposed method uses the training method to obtain the relevant parameters.

4. Medical Image Classification Algorithm Based on Weight Initialization-Sliding Window Fusion Convolutional Neural Network

Based on the above, this section builds a medical image classification algorithm based on a weight initialization/sliding window fusion CNN. First, a CNN weight initialization method is established, which improves the convergence and generalizability of the model and avoids problems such as gradients in the CNN that occur due to weight initialization problems. Then, the weight initialization method is introduced into the sliding window fusion CNN model proposed in the second part of this paper to improve the adaptive ability of the multilayer CNN sliding window fusion model. Finally, a medical image classification algorithm based on a weight initialization/sliding window fusion CNN is established. The basic flow chart of the proposed medical image classification algorithm is shown in Figure 3. The basic steps are as follows:

- (1) First, the medical image is preprocessed (e.g., denoised).
- (2) Weighted neural network weight initialization is performed using the adaptive Taylor weight initialization method proposed in this paper. Compared with other methods, this method has strong versatility and can yield analytical solutions. It also improves the adaptive ability and feature extraction ability of the CNN model; thus, it extracts more medical image feature information.
- (3) To improve the complex feature information of medical images that cannot be completely obtained by a single CNN, this paper differentiates the characteristics of the same medical image by constructing different network structures, thereby improving the generalization performance of the entire network. Moreover, for the fusion of multilayer networks, a sliding window fusion mechanism is proposed to realize flexible selection of multilayer network classification results. It also optimizes the

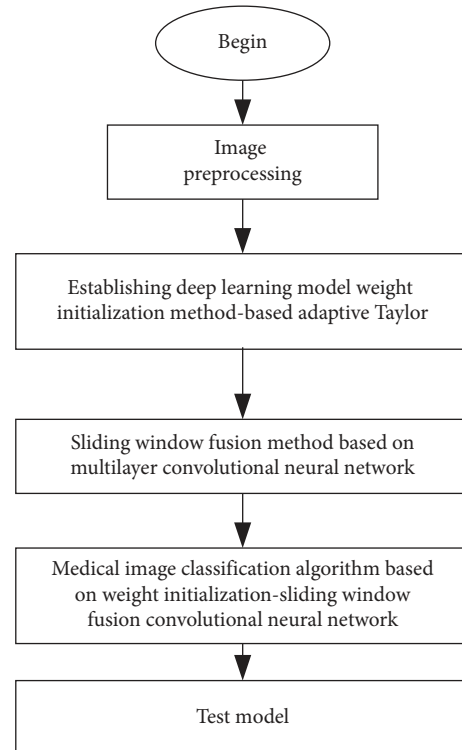


FIGURE 3: Basic flow chart of the medical classification algorithm proposed in this paper.

fusion process of multilayer CNNs and improves the accuracy of medical image classification.

- (4) The methods of step (2) to step (3) are combined, and a medical image classification algorithm based on weight initialization-sliding window fusion convolutional neural network is established through steps (1)–(3). The medical image classification algorithm is used to analyze related examples and obtain the classification results.

5. Experiment Analysis

5.1. Breast Mass Classification Experiment. The breast mass image dataset used in this experiment is the Digital Database for Screening Mammography (DDSM) released by the University of South Florida. The dataset includes labels that indicate benign and malignant breast masses and lesion-level annotated information for pixel-level accuracy. The specific experimental data were set as follows: 600 images from the dataset (300 benign mass images and 300 malignant mass images) were divided into training sets and test sets according to a 60 : 40 ratio, respectively, with equal numbers of benign and malignant masses in each set.

To verify the effectiveness of the proposed method, this paper selects the average classification accuracy of multiple randomized datasets (the average of 100 randomized partitions) as the actual test results and adopts the mainstream methods ([48] refers to the machine learning classification method and [49, 50] refer to the optimized CNN methods) for comparison. The specific test results are shown in Table 2.

TABLE 2: Comparison table of breast tumor image classification results.

Classification method	Classification accuracy (%)
[48]	95.4
CNN	96.5
[49]	96.7
[50]	97.1
Method of this paper	98.2

Table 2 shows that the medical image classification method proposed in this paper achieves better results than the other methods. The accuracy of the proposed method is 3.8 percentage points higher than that in [48] and 1.7 percentage points higher than that of the traditional CNN. This finding shows that the proposed method has greater advantages over the machine learning method and the traditional CNN. The classification accuracy obtained by the proposed method is 1.5 percentage points and 1.1 percentage points higher than that of the methods in [49, 50], respectively. This finding further verifies that the proposed method is superior to other mainstream CNN-based medical image classification methods. To better explain why the proposed method is superior to other CNN-based methods, the following explains the theory behind and the practical calculations of the method.

In this paper, for the image of a breast mass, the gray features of the original breast mass image, the features of the CNN feature layer, and the CNN features of the feature transformation were visualized. The results are shown in Figure 4. Figures 4(a)–4(c) represent the original grayscale features of the breast mass image, the feature vectors of the network layer, and the visualization results of the network layer features processed by the method, respectively. In each figure, the yellow and blue dots represent benign mass image samples and malignant mass image samples, respectively. By comparing the visualization results of the different layer feature distributions in the figure, it can be known that the CNN features and the features after the multislicing window transformation greatly improve discrimination between the original benign and malignant breast mass images. In addition, the CNN features are represented by the network layer without the large-interval metric-learning layer transformation. The large-interval metric-learning layer can transform the original breast mass image CNN features into a new feature space with a more compact distribution and a more discrete distribution between classes. It makes the different types of masses more distinguishable in the feature space, thereby further improving the classification accuracy.

5.2. Brain Tumor Tissue Classification Experiment. The experimental data are from the Third Hospital of Peking University. The dataset contains cross-sectional MRI images of the brains of 12 patients with brain tumors. All the images were preprocessed in terms of format, noise reduction, etc., and the neurosurgical and imaging surgeons mapped the true area of the brain tumor as a training sample label and a reference standard for the comparison experiments.

Moreover, a large number of samples are required for training CNN models. To expand the number of training samples and test the results of this method to classify different brain tumor images, the experiment used ten-fold cross-validation by the leave-one-out method. For the 12 MRI images of brain tumors, 11 images were selected for the training set, and 1 image was used as the test set. The numbers of training samples and test samples collected in each experiment are shown in Table 3.

In the experiment, the classification method proposed in this paper was tested first. To verify the effectiveness of the proposed method, it was compared with traditional machine learning methods and other deep learning methods. All the experimental data and the experimental environments were consistent, and the relevant parameter settings were obtained by the optimization method. The brain tumor tissue classification results are shown in Table 4.

As shown in Table 4, the average accuracy for the traditional machine learning method (i.e., the support vector machine (SVM)) is only 88.19% because the SVM method is a small sample method; it is not as effective for large samples as deep learning methods. The average accuracy of the traditional CNN method is 91.12%—nearly 3 percentage points higher than that of the SVM method. This result shows that the deep learning method has large advantages over the traditional machine learning method. The average classification accuracy of the optimized CNN method proposed in [2] is 93.67%, which is not only much higher than the traditional SVM method but also higher than the traditional CNN method because the method suppresses the overfitting problem of the CNN to a certain extent and consequently achieves a good classification effect. The average accuracy of the method proposed in this paper reached 95.71% on the image block classification task of the MRI images of 12 patients with brain tumors, which was the highest accuracy among the tested models, especially on the 2nd, 11th, and 12th samples, which were more difficult to classify than the other samples. This experiment fully demonstrates that the proposed method not only greatly improves the classification accuracy but also has good stability and robustness and better solves the problems of weight initialization and overfitting in CNNs. Thus, the model structure can take full advantage of the deep learning method.

5.3. Medical Database Classification Experiment. To further validate the classification effect of the proposed algorithm on medical images, this section conducts classification tests on two public medical databases (i.e., The Cancer Imaging Archive (TCIA)-CT database [51] and the Open Access Series of Imaging Studies (OASIS)-MRI database [52]) and mainstream images. The classification algorithms were evaluated through a comparative analysis. This study adopted the same settings for the TCIA-CT database as those in [53]; the images are of the Digital Imaging and Communications in Medicine (DICOM) type. This analysis was used for the experimental tests in this section, and examples are shown in Figure 5. This study selected 604 colon images

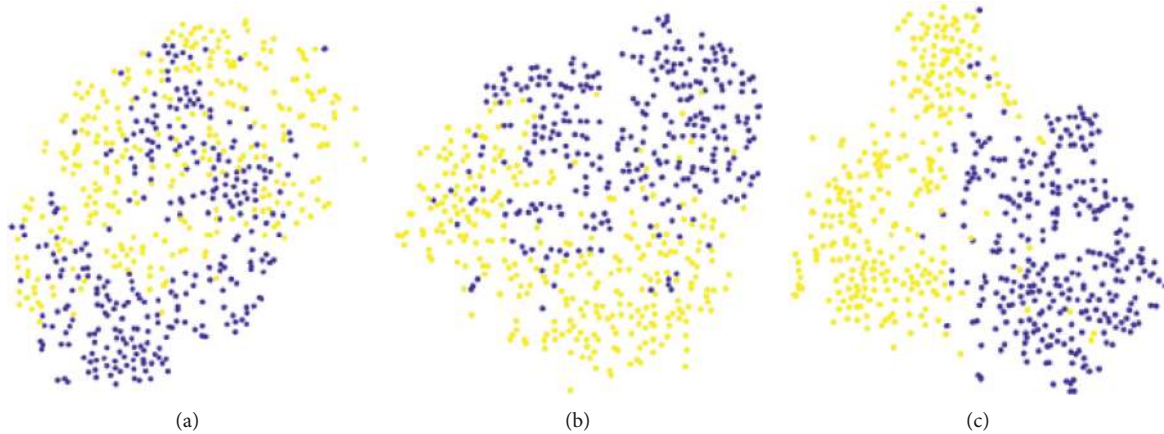


FIGURE 4: Visual distributions of the breast mass features: (a) original underlying features; (b) network layer features; (c) features learned by the proposed method.

TABLE 3: Samples for each experiment.

Rounds	Training samples	Test sample
1	150380	76502
2	160920	76530
3	165400	79980
4	177000	80640
5	150820	71600
6	135340	68850
7	144828	68877
8	148860	71980
9	159300	72576
10	135738	64440
11	181940	87978
12	194700	88704

TABLE 4: Brain tumor tissue classification results.

Sample number	SVM (%)	CNN (%)	[2] (%)	Our (%)
1	89.41	91.09	92.82	94.70
2	76.71	85.26	89.76	93.48
3	93.32	94.32	95.15	97.12
4	95.29	96.01	97.89	98.88
5	94.31	95.32	97.18	99.16
6	94.87	96.02	97.90	98.89
7	91.05	92.16	93.92	95.82
8	89.87	91.09	93.82	95.70
9	86.54	89.05	91.72	93.55
10	89.33	92.03	93.79	95.69
11	78.23	86.05	90.57	93.30
12	79.31	85.06	89.55	92.26
Average value	88.19	91.12	93.67	95.71

from the database and used the data enhancement strategy to improve the database to obtain a training dataset of 988 images and a test dataset of 218 images. The OASIS-MRI database was divided into four categories, each of which contains 152, 121, 88, and 68 images; examples are shown in Figure 6. This image set was processed using the same data selection and data enhancement method as was used on the TCIA-CT database, generating a training set of 498 images and a test set of 86 images.

The two abovementioned medical image databases were classified by the classification algorithm proposed in this paper and other mainstream image classification algorithms, and the results are shown in Table 5.

It can be seen from Table 5 that the medical image classification algorithm proposed in this paper has better classification effect than traditional medical classification algorithms and other deep learning medical classification algorithms. Moreover, the proposed method performs stably on both medical image databases. Specifically, the first three traditional classification algorithms in the table are mainly to separate image feature extraction and classification into two steps and then combine them to classify the medical images. The three corresponding deep learning algorithms in the table unify the feature extraction and classification process into a single task to complete the corresponding test. Therefore, the robustness and accuracy of the classification results obtained by the integrated classification algorithms are higher than those of the combined traditional methods.

Specifically, for the performance in the TCIA-CT database, only the algorithm proposed in this paper achieves the best classification results. The three traditional medical image classification algorithms are less accurate. The results of the two deep learning models DeepNet1 and DeepNet3 are satisfactory; they have certain advantages over traditional methods. For the more difficult to classify OASIS-MRI database, all the deep models perform significantly better than the traditional machine learning algorithms. This result indicates that the accuracy of automatically learning deep features when applied to medical image classification tasks is much higher than the accuracy of artificially designed image features.

In short, the traditional classification algorithms have the disadvantages of low classification accuracy and poor stability on medical image classification tasks. It shows that this combined traditional classification method does not work well for biomedical image classification. The algorithms' classification accuracy of the deep learning algorithms on the two medical image databases is significantly better than those of the traditional classification algorithms, again

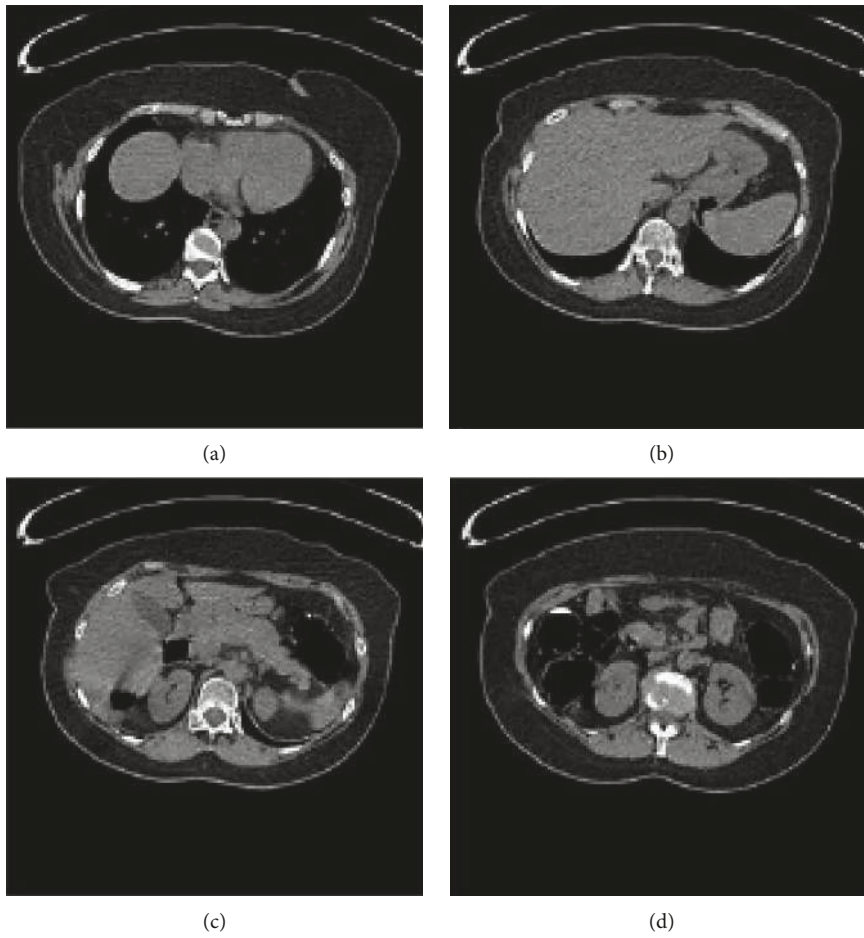


FIGURE 5: Examples from the TCIA-CT database.

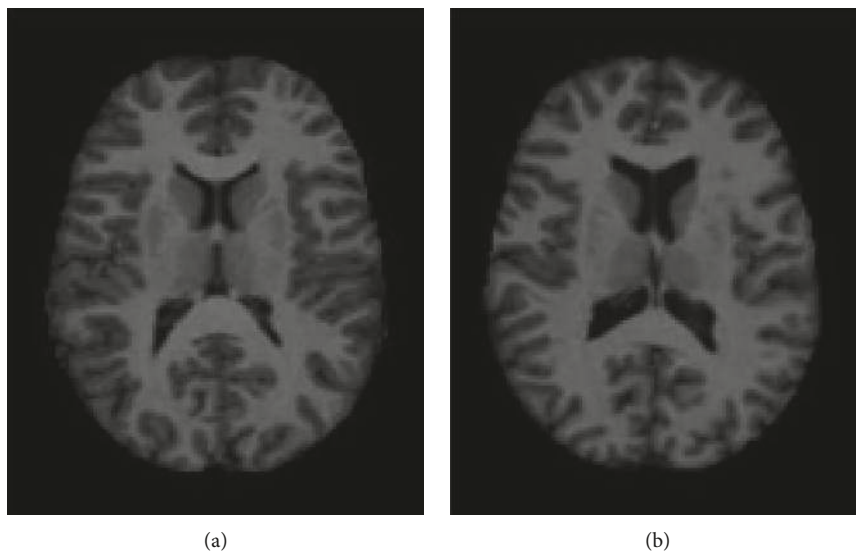


FIGURE 6: Continued.

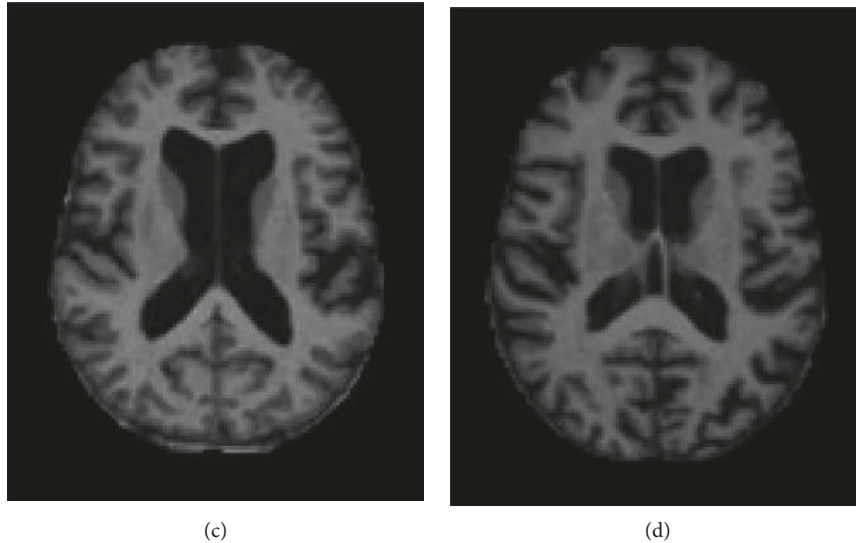


FIGURE 6: Examples from the OASIS-MRI database.

TABLE 5: Comparison of classification accuracy of different classification algorithms on two medical image databases (unit: %).

Method type	Medical database type	
	TCIA-CT	OASIS-MRI
LBP + SVM	71.8	57.5
HOG + KNN	85.1	67.6
HOG + SVM	87.3	81.6
DeepNet1	98.7	89.2
DeepNet3	99.2	92.1
Our	100	95.9

indicating that deep learning models have advantages from the side. In addition, the deep learning model medical image classification algorithms have high stability. In particular, the medical image classification algorithm proposed in this paper achieves better effects than the other deep learning medical classification algorithms. This result occurs because the deep learning model proposed in this paper not only solves the problem of model weight initialization but also solves the problem of multilayer association in the deep learning model.

5.4. Brain Medical Database Classification Experiment. In order to further verify that the proposed algorithm can have a good classification effect on general medical databases, this paper randomly downloads 395 participants from the public dataset Alzheimer’s Disease Neuroimaging Initiative (ADNI) [54]. These participants were composed of 101 patients with early Alzheimer’s disease (AD), 145 patients with mild impairment cognitive (MCI), and 149 normal control (NC). All longitudinal MRI scan images are obtained using a professional scanner. Each participant has 1 to 10 scanned images; that is, participants perform MRI scans at up to 10 different time points. Moreover, each scanned image has an average of 180 slice images. Part of the image is shown in Figure 7. At the same time, in order to better adapt to the composition of relevant datasets, the paper randomly

selected the corresponding participants from the participants of AD, MCI, and NC according to the ratio of 35%, 30%, and 35% and obtained 9 sets of data. Among them, 35% of the selected participants merged and got the training set. The other 35% of the selected participants merged and trained to get the test set. The remaining 30% of the selected participants are merged and trained to obtain a validation set.

The classification algorithm and the other mainstream image classification algorithms are used to classify the medical image databases randomly selected by the ADNI database. The classification results are shown in Table 6.

As can be seen from Table 6, the traditional machine learning method as in reference [55] is only 88.3%. The brain medical database constructed in this paper is composed of randomly selected data in the ADNI database, which has greater randomness and uncertainty. Traditional machine learning methods cannot extract more effective feature information from these random images. The classification accuracy of the CNN algorithm is only 0.9% higher than that of the traditional machine learning algorithm. The main factor may be that the CNN network structure is not well adapted to the image characteristics in the brain medical database built in this paper. Other deep learning algorithms such as reference [56–60] have classification accuracy of 90.3%, 91.2%, 92.3%, 92.3%, and 92.5%, respectively. It can be seen that the classification accuracy of these deep learning algorithms is above 90%, which also shows that the deep learning algorithm constructed by the subsequent use of brain medical image feature information can better utilize brain medical image information for classification. The classification accuracy obtained by the proposed method reaches 95.1%, which is 2% higher than other deep learning algorithms. This is mainly because this paper proposes a network weight initialization method that can better solve the deep learning model from the network structure of deep learning. It also increases the potential of convolutional neural networks to handle different visual tasks. At the same time, the framework of multicolumn convolutional neural

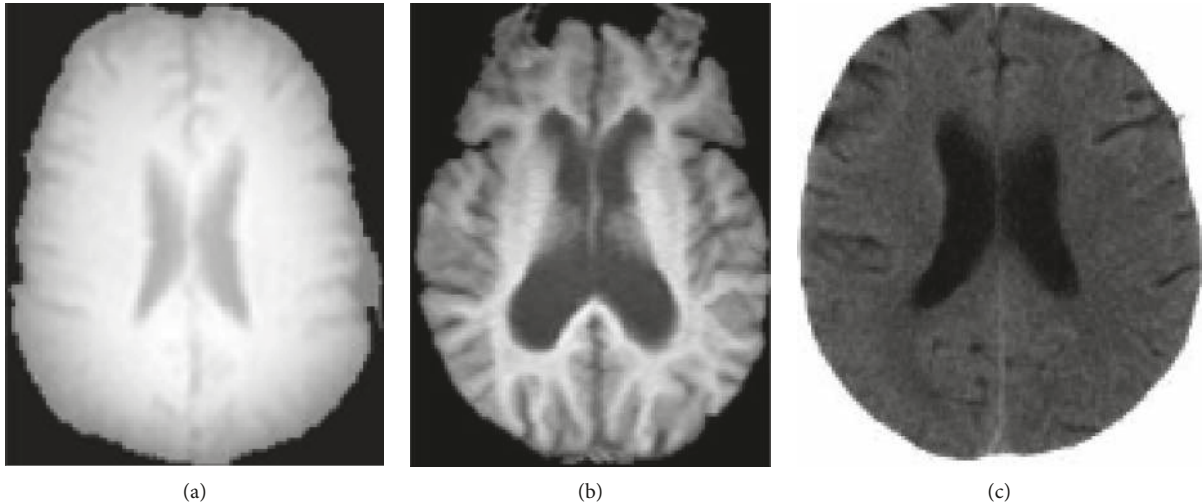


FIGURE 7: Examples from the ADNI database.

TABLE 6: Brain medical image classification results comparison table.

Method type	Classification accuracy (%)
Zhang et al. [55]	88.3
CNN	89.2
Suk et al. [56]	90.3
Liu et al. [57]	91.2
Liu et al. [58]	92.3
Shen et al. [59]	92.3
Aderghal et al. [60]	92.5
Our	95.1

network is constructed according to the characteristics of medical images, and the sliding window fusion processing is performed on the multicolumn convolutional neural network. It can obtain a better convolutional neural network model and achieve high-precision classification of brain medical databases. The experiment and the above experiments also show that the medical image classification algorithm proposed in this paper can classify many different medical images well, which further validates the validity and reliability of the proposed algorithm. At the same time, it further confirms that the proposed method has good stability and robustness.

6. Conclusion

From the network optimization perspective, this paper uses optimization methods to improve the nonlinear modeling ability of the network. By improving the attracting domain to improve the generalization performance of the solution after convergence, a new method of network weight initialization is proposed. This method alleviates the problem that the initialization theory of existing deep learning is limited by the type, and it increases the potential of the neural network to address different visual tasks. Moreover, this paper constructs a CNN model with a variety of structures that better adapt

to the characteristics of the medical images of interest. The different CNN models are trained on the same dataset, and finally, the trained heterogeneous multicolumn CNN is combined with the adaptive sliding window fusion mechanism proposed in this paper. Together, they accomplish the task of classifying medical images.

The results of the brain tumor tissue classification experiments show that the proposed method has the highest average accuracy, reaching 95.71%. On the image block classification task involving MRI samples of 12 patients with brain tumors, the proposed model's advantages were clearly shown on samples 2, 11, and 12, which are more difficult to classify than the other samples. This experiment shows that the method proposed in this paper has good stability and robustness while greatly improving the classification accuracy.

The results of the medical database classification experiments show that the classification algorithm proposed in this paper achieves the best classification results for the TCIA-CT database and for the images in the OASIS-MRI database that are the most difficult to classify. The proposed method is not only superior to traditional machine learning algorithms but also improves other deep learning models.

The results of brain-based self-built medical database classification experiments show that the proposed method has the highest accuracy rate of 95.1% on the brain medical database constructed on the basis of ADNI database. It further verifies that the proposed method can better adapt to the classification task of general medical database.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This paper was supported by the National Natural Science Foundation of China (no. 61701188) and China Postdoctoral Science Foundation (no. 2019M650512). This work was supported in part by National Natural Science Foundation of China (no. 61701188), Postdoctoral Science Foundation of China (no. 2019M650512), and Natural Science Foundation of Shanxi (no. 201801D221171).

References

- [1] Y. Song, W. Cai, H. Huang et al., "Large margin local estimate with applications to medical image classification," *IEEE Transactions on Medical Imaging*, vol. 34, no. 6, pp. 1362–1377, 2015.
- [2] H. Mohsen, E.-S. A. El-Dahshan, E.-S. M. El-Horbaty, and A.-B. M. Salem, "Classification using deep learning neural networks for brain tumors," *Future Computing and Informatics Journal*, vol. 3, no. 1, pp. 68–71, 2018.
- [3] S. K. Zhou, H. Greenspan, and D. Shen, *Deep Learning for Medical Image Analysis*, Academic Press, Cambridge, MA, USA, 2017.
- [4] J. Verma, M. Nath, P. Tripathi, and K. K. Saini, "Analysis and identification of kidney stone using K^{th} nearest neighbour (KNN) and support vector machine (SVM) classification techniques," *Pattern Recognition and Image Analysis*, vol. 27, no. 3, pp. 574–580, 2017.
- [5] F. A. Spanhol, L. S. Oliveira, P. R. Cavalin, C. Petitjean, and L. Heutte, "Deep features for breast cancer histopathological image classification," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1868–1873, Banff, Canada, October 2017.
- [6] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1455–1462, 2016.
- [7] A. Esteva, B. Kuprel, R. A. Novoa et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [8] D. Liu, S. Wang, D. Huang, G. Deng, F. Zeng, and H. Chen, "Medical image classification using spatial adjacent histogram based on adaptive local binary patterns," *Computers in Biology and Medicine*, vol. 72, pp. 185–200, 2016.
- [9] G. Mohan and M. M. Subashini, "MRI based medical image analysis: survey on brain tumor grade classification," *Biomedical Signal Processing and Control*, vol. 39, pp. 139–161, 2018.
- [10] J. Wu, Y. Yu, and C. Huang, "Deep multiple instance learning for image classification and auto-annotation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3460–3469, Boston, MA, USA, June 2015.
- [11] R. J. Ramteke and Y. K. Monali, "Automatic medical image classification and abnormality detection using k -nearest neighbor," *International Journal of Advanced Computer Research*, vol. 2, no. 4, pp. 190–196, 2012.
- [12] J. J. Corso, E. Sharon, S. Dube, S. El-Saden, U. Sinha, and A. Yuille, "Efficient multilevel brain tumor segmentation with integrated Bayesian model classification," *IEEE Transactions on Medical Imaging*, vol. 27, no. 5, pp. 629–640, 2008.
- [13] A. Rao, Y. Lee, A. Gass, and A. Monsch, "Classification of Alzheimer's disease from structural MRI using sparse logistic regression with optional spatial regularization," in *Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4499–4502, Boston, MA, USA, August 2011.
- [14] S. Issac Niwas, P. Palanisamy, and K. Sujathan, "Complex wavelet as nucleus descriptors for automated cancer cytology classifier system using ANN," in *Proceedings of the IEEE 2010 International Conference on Computational Intelligence and Computing Research*, pp. 1–5, Coimbatore, India, December 2010.
- [15] Z. Camlica, H. R. Tizhoosh, and F. Khalvati, "Medical image classification via SVM using LBP features from saliency-based folded data," in *Proceedings of the 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pp. 128–132, Miami, FL, USA, December 2015.
- [16] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Proceedings of the Workshop on Statistical Learning in Computer Vision (ECCV)*, pp. 1–4, Prague, Czech Republic, May 2004.
- [17] A. Singh, "Detection of brain tumor in MRI images, using combination of fuzzy c-means and SVM," in *Proceedings of the 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN)*, pp. 98–102, New Delhi, India, February 2015.
- [18] Y.-D. Zhang, S. Chen, S.-H. Wang, J.-F. Yang, and P. Phillips, "Magnetic resonance brain image classification based on weighted-type fractional fourier transform and nonparallel support vector machine," *International Journal of Imaging Systems and Technology*, vol. 25, no. 4, pp. 317–327, 2015.
- [19] S. Zhu and H. R. Tizhoosh, "Radon features and barcodes for medical image retrieval via SVM," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, pp. 5065–5071, Vancouver, Canada, July 2016.
- [20] H. A. Abbass, "An evolutionary artificial neural networks approach for breast cancer diagnosis," *Artificial Intelligence in Medicine*, vol. 25, no. 3, pp. 265–281, 2002.
- [21] M. Karabatak and M. C. Ince, "An expert system for detection of breast cancer based on association rules and neural network," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3465–3469, 2009.
- [22] L.-H. Juang and M.-N. Wu, "MRI brain lesion image detection based on color-converted K-means clustering segmentation," *Measurement*, vol. 43, no. 7, pp. 941–949, 2010.
- [23] D.-Q. Zhang and S.-C. Chen, "A novel kernelized fuzzy C-means algorithm with application in medical image segmentation," *Artificial Intelligence in Medicine*, vol. 32, no. 1, pp. 37–50, 2004.
- [24] D. Singh and K. Kaur, "Classification of abnormalities in brain MRI images using GLCM, PCA and SVM," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 1, no. 6, pp. 243–248, 2012.
- [25] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [26] F. Radenović, G. Toliás, and O. Chum, "CNN image retrieval learns from BoW: unsupervised fine-tuning with hard examples," in *Proceedings of the European Conference on Computer Vision*, pp. 3–20, Amsterdam, Netherlands, October 2016.
- [27] V. Campos, B. Jou, and X. Giró-i-Nieto, "From pixels to sentiment: fine-tuning CNNs for visual sentiment prediction," *Image and Vision Computing*, vol. 65, pp. 15–22, 2017.
- [28] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440–1448, Santiago, Chile, December 2015.

- [29] T. Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear CNN models for fine-grained visual recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1449–1457, Santiago, Chile, December 2015.
- [30] T. Brosch and R. Tam, "Manifold learning of brain MRIs by deep learning," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 633–640, Nagoya, Japan, September 2013.
- [31] S. M. Plis, D. R. Hjelm, R. Salakhutdinov et al., "Deep learning for neuroimaging: a validation study," *Frontiers in Neuroscience*, vol. 8, pp. 229–238, 2014.
- [32] J. Z. Cheng, D. Ni, Y. H. Chou et al., "Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans," *Scientific Reports*, vol. 6, no. 1, pp. 24454–24466, 2016.
- [33] M. Kallenberg, K. Petersen, M. Nielsen et al., "Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1322–1331, 2016.
- [34] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu et al., "Convolutional neural networks for medical image analysis: full training or fine tuning?," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.
- [35] Z. Liang, A. Powell, and I. Ersoy, "CNN-based image analysis for malaria diagnosis," in *Proceedings of the 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 493–496, Shenzhen, China, December 2016.
- [36] M. Gao, U. Bagci, L. Lu et al., "Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 6, no. 1, pp. 1–6, 2018.
- [37] C. K. Shie, C. H. Chuang, C. N. Chou, M.-H. Wu, and E. Y. Chang, "Transfer representation learning for medical image analysis," in *Proceedings of the 2015 37th annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 711–714, Milan, Italy, August 2015.
- [38] Q. Zhu, B. Du, B. Turkbey, P. L. Choyke, and P. Yan, "Deeply-supervised CNN for prostate segmentation," in *Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 178–184, Anchorage, AK, USA, May 2017.
- [39] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification," in *Proceedings of the International Conference on Information Processing in Medical Imaging*, pp. 588–599, Isle of Skye, Scotland, July 2015.
- [40] J. Kawahara and G. Hamarneh, "Multi-resolution-tract CNN with hybrid pretrained and skin-lesion trained layers," in *Proceedings of the International Workshop on Machine Learning in Medical Imaging*, pp. 164–171, Athens, Greece, October 2016.
- [41] R. Bidart, M. J. Gangeh, and M. Peikari, "Localization and classification of cell nuclei in post-neoadjuvant breast cancer surgical specimen using fully convolutional networks," in *Proceedings of the Medical Imaging 2018: Digital Pathology*, pp. 1–8, Houston, TX, USA, February 2018.
- [42] K. He, X. Zhang, S. Ren, and S. Jian, "Delving deep into rectifiers: surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1026–1034, Santiago, Chile, December 2015.
- [43] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 249–256, Sardinia, Italy, May 2010.
- [44] G. Klambauer, T. Unterthiner, and A. Mayr, "Self-normalizing neural networks," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 971–980, Long Beach, CA, USA, December 2017.
- [45] M. Jaderberg, K. Simonyan, and A. Zisserman, "Spatial transformer networks," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 2017–2025, Montreal, Canada, December 2015.
- [46] X. Wang, R. Girshick, and A. Gupta, "Non-local neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7794–7803, Salt Lake City, UT, USA, June 2018.
- [47] C. Gulcehre, M. Moczulski, and M. Denil, "Noisy activation functions," in *Proceedings of the International Conference on Machine Learning*, pp. 3059–3068, New York, NY, USA, June 2016.
- [48] W. Xie, Y. Li, and Y. Ma, "Breast mass classification in digital mammography based on extreme learning machine," *Neurocomputing*, vol. 173, pp. 930–941, 2016.
- [49] Y. Zheng, J. Fan, J. Zhang, and X. Gao, "Hierarchical learning of multi-task sparse metrics for large-scale image classification," *Pattern Recognition*, vol. 67, pp. 97–109, 2017.
- [50] Y. Qiu, S. Yan, R. R. Gundreddy, Y. Wang, S. Cheng, H. Liu et al., "A new approach to develop computer-aided diagnosis scheme of breast mass classification using deep learning technology," *Journal of X-Ray Science and Technology*, vol. 25, no. 5, pp. 751–763, 2017.
- [51] K. Clark, B. Vendt, K. Smith et al., "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.
- [52] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults," *Journal of Cognitive Neuroscience*, vol. 19, no. 9, pp. 1498–1507, 2007.
- [53] S. R. Dubey, S. K. Singh, and R. K. Singh, "Local wavelet pattern: a new feature descriptor for image retrieval in medical CT databases," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5892–5903, 2015.
- [54] <http://adni.loni.usc.edu/>.
- [55] J. Zhang, M. Liu, L. Le An, Y. Gao, and D. Shen, "Alzheimer's disease diagnosis using landmark-based features from longitudinal structural MR images," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 6, pp. 1607–1616, 2017.
- [56] H.-I. Suk, S.-W. Lee, and D. Shen, "Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis," *NeuroImage*, vol. 101, pp. 569–582, 2014.
- [57] M. Liu, D. Cheng, and W. Yan, "Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images," *Frontiers in Neuroinformatics*, vol. 12, pp. 35–43, 2018.
- [58] M. Liu, D. Cheng, K. Wang, and Y. Wang, "Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis," *Neuroinformatics*, vol. 16, no. 3–4, pp. 295–308, 2018.
- [59] T. Shen, J. Jiang, Y. Li, P. Wu, C. Zuo, and Z. Yan, "Decision supporting model for one-year conversion probability from MCI to AD using CNN and SVM," in *Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering*

in Medicine and Biology Society (EMBC), pp. 738–741, Honolulu, HI, USA, July 2018.

- [60] K. Aderghal, A. Khvostikov, A. Krylov, J. Benois-Pineau, K. Afdel, and G. Catheline, “Classification of Alzheimer disease on imaging modalities with deep CNNs using cross-modal transfer learning,” in *Proceedings of the 2018 IEEE 31st International Symposium on Computer-Based Medical Systems*, pp. 345–350, Karlstad, Sweden, June 2018.

