A Comprehensive Review of Medical Image Analysis Technology Based on CNN

Zixuan Wang*

School of Mathematical Sciences, Soochow University, Suzhou, China.

2207402094@stu.suda.edu.cn

Abstract. In recent years, artificial intelligence technology has achieved remarkable breakthroughs in the field of healthcare. CNN, with its powerful image feature extraction capability, has become a vital technical pillar in the field of medical image analysis. At the model level, this paper systematically combs the innovative characteristics of typical CNN architectures and their adaptive improvements in the medical imaging field, clarifying the technological development context. At the application level, it focuses on four core tasks of medical image analysis, i.e., classification, boundary detection, registration, and text-integrated analysis. Datasets corresponding to each task and their acquisition methods are also meticulously organized, providing technical guidance and data support for related research.

Keywords: CNN; Medical Image; Medical Image Analysis.

1. Research Background

In recent years, artificial intelligence technology has demonstrated revolutionary potential in the healthcare field. Among them, machine learning significantly improves the automation level of medical image analysis by automatically extracting multi-level features from data to establish prediction models[1]. As an important branch of machine learning, deep learning can mine potential pathological feature representations from massive medical imaging data by constructing deep neural networks with multiple nonlinear transformations. Among numerous deep learning architectures, the Convolutional Neural Network (CNN) has become the most influential technical framework in medical image analysis due to its unique local connection and weight sharing mechanisms[2]. This paper focuses on the technological development of CNN and its medical applications. At the model level, CNN has undergone continuous evolution from early classic architectures (such as LeNet and AlexNet) to modern deep networks (such as ResNet and EfficientNet), with core innovations embodied in network depth optimization, attention mechanism integration, and lightweight design[3]. This paper systematically combs the architectural characteristics of these key models and their adaptive improvements for medical image analysis, providing readers with a clear technological development context.In terms of medical applications, CNN research mainly revolves around four core tasks:

(1) Image classification: such as chest X-ray disease classification based on the CheXpert dataset, focusing on transfer learning strategies (e.g., DenseNet-121)[4];

(2) Lesion detection: such as lung nodule detection using the LUNA16 dataset, discussing optimization methods for frameworks like Faster R-CNN and YOLO[5];

(3) Anatomical structure segmentation: such as brain tumor segmentation based on the BraTS dataset, studying performance of models like nnUNet in multi-modal MRI[6];

(4) Image registration: such as CT registration for DIRLAB dataset, analyzing balance between accuracy and efficiency of unsupervised methods like VoxelMorph[7].

For each type of task, this paper not only summarizes representative algorithms but also conducts detailed comparisons of dataset types, scale characteristics, and their impacts on model training. Medical image analysis faces unique challenges such as small-sample learning, class imbalance, and annotation noise, making the rational selection and utilization of datasets crucial to solving these problems[5].

2. The Birth and Basic Principles of CNN

Following the structure and working principles of biological neurons, McCulloch and Pitts proposed a simplified McCulloch-Pitts (M-P) neuron model in 1943[8]. This model, the basic element of neural networks, paved the way for the development of CNN. Based on the basic neuron, the LeNet and LeNet-5 models[9] were born, marking the beginning of CNN models. The AlexNet[10] and GoogLeNet[11] models became milestones in CNN history by continuously optimizing CNN[12].

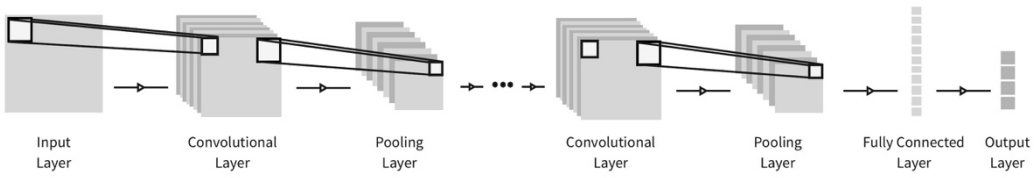

Fig. 1 The Basic Structure of CNN

2.1 Convolutional Layer

The convolutional layer in CNN achieves feature extraction through the mechanisms of local connection and weight sharing, and its core operation can be described using strict mathematical formulas. For a two-dimensional input feature map $X \in \mathbb{R}^{\mathrm{H_{in}}\times\mathrm{W_{in}}\times\mathrm{D_{in}}}$ and a convolution kernel $K \in \mathbb{R}^{k_\mathrm{h}\times k_\mathrm{w}\times D_\mathrm{in}\times D_{out}}$, the convolution operation can be expressed as:

$$Y(i,j,d) = \sum_{c=0}^{D_\mathrm{in}-1}\sum_{m=0}^{k_h-1}\sum_{n=0}^{k_w-1} X(s\cdot i+m-p, s\cdot j+n-p, c)\cdot K(m,n,c,d) + b_d,$$

Here, $Y \in \mathbb{R}^{H_\mathrm{out}\times W_\mathrm{out}\times D_\mathrm{out}}$ is the output feature map, s is the stride, p is the padding, and b_d is the bias term for the d-th output channel[9]. The size of the output feature map is jointly determined by the input size (H_{in}, W_{in}), the convolution kernel size (k_h, k_w), the stride s, and the padding p. The calculation formula[9] is:

$$H_\mathrm{out} = \left\lfloor \frac{H_\mathrm{in}+2p-k_h}{s} \right\rfloor + 1 \quad W_\mathrm{out} = \left\lfloor \frac{W_\mathrm{in}+2p-k_w}{s} \right\rfloor + 1.$$

The convolution kernel size (k_h, k_w) determines the size of the local receptive field. A larger convolution kernel can capture broader regional features but will increase the computational load. The stride s controls the sliding interval of the convolution kernel on the input feature map. When the stride is greater than 1, it will produce a downsampling effect, significantly reducing the spatial size of the output feature map. To maintain the consistency of the input and output sizes, a padding operation is usually adopted. In the "valid" mode, no padding is performed, while in the "same" mode, symmetric padding is used to make the output size the same as the input size. At this time, the padding amount needs to satisfy $p = \lfloor (k_h-1)/2 \rfloor$ [13].

The Receptive Field (RF) is an important concept in the CNN, representing the input region range corresponding to each neuron on the output feature map. RF_l of the l-th layer can be calculated by:

$$RF_l = RF_{l-1} + \left(k_h^{(l)} - 1\right) \times \prod_{i=1}^{l-1} s_i,$$

Among them, $k_h^{(l)}$ is the height of the convolution kernel in the l-th layer, and s_i is the stride of the i-th layer[9]. As the network depth increases, the receptive field of deep-layer neurons will expand exponentially, enabling high-layer neurons to capture semantic information in a larger range of the input image.

The convolution depth D_out is determined by the number of convolution kernels. Each convolution kernel will perform an independent convolution operation on the input feature map to generate one channel of the output feature map. In the case of multi-channel input, the depth dimension D_in of each convolution kernel must match the number of channels of the input feature

map. By performing convolution operations on all input channels and summing up the results, a single-channel output feature map is finally generated. By adjusting these parameters, the structure of the convolutional layer can be flexibly designed to achieve a balance between computational efficiency and feature expression ability[14].

2.2 Pooling Layer

The pooling layer is an important component in CNN. Its main function is to reduce the spatial size of the feature map through down-sampling, thereby reducing the computational complexity and enhancing the robustness of the model. The pooling layer realizes the compression of spatial dimensions by performing aggregation statistics on the local regions of the input feature map. Common pooling operations include Max pooling[9] and Average pooling[15]. The pooling layer plays multiple crucial roles in a CNN, and its core values are mainly reflected in the following aspects. First, the pooling layer significantly improves computational efficiency through dimensionality reduction. Second, the pooling layer, especially max-pooling, can endow the model with good translation invariance. By retaining the most prominent features in the local region, the network becomes more robust to geometric transformations such as small translations and rotations of the input data. Moreover, the pooling operation also has a certain regularization effect. Since it reduces the dimension of intermediate features, it objectively suppresses overfitting and enhances the generalization ability of the model. In addition, as the network depth increases, the stacked use of pooling layers can gradually expand the receptive field of high-level neurons, enabling them to cover a larger range of the input image, thus helping to capture more global semantic information. It is worth noting that although the pooling operation will lose some spatial detail information, this selective discarding of information instead makes the network pay more attention to discriminative high-level features. In modern network architectures, although there is a trend of replacing traditional pooling layers with operations like strided convolutions, the pooling layer, due to its simplicity and efficiency, remains an indispensable component in many classic models, especially when balancing computational efficiency and feature robustness[16].

2.3 Fully Connected (FC) and Output Layers

The FC layer is a core component in a neural network. Its role is to fully connect all neurons of the previous layer with each neuron of the current layer, achieving global integration of features and high-order nonlinear transformations. The mathematical expression is $y = \sigma(Wx + b)$,where $x \in \mathbb{R}^n$ is the input feature vector,$W \in \mathbb{R}^{m \times n}$ is the weight matrix, $b \in \mathbb{R}^m$ is the bias vector, and $\sigma(\cdot)$ is the activation function (such as ReLU, Sigmoid). This layer integrates features from all spatial positions and channels of the previous layer through dense connections, and is suitable for the decision-making stage of classification tasks. The output dimension can be flexibly adjusted to achieve dimension transformation of the feature space. However, when the input size is too large, it is prone to overfitting.

The design of output layer depends on the type of task. For classification tasks, the Softmax output is suitable for multi-classification, outputting the probability distribution of classes, where K is the number of classes[17]:

$$p_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}, i = 1, \ldots, K.$$

The Sigmoid output is suitable for outputting the probability of a single node in binary classification tasks.For regression tasks, the linear output directly outputs continuous values without an activation function $y = w^T x + b$.[17]

The FC layer maps the local features extracted by the convolutional layer to a global representation, and the output layer converts it into the prediction form required by the task. Both participate in backpropagation, and through the chain rule, they achieve the gradient transfer from the loss function to the low-level features.

Table 1 Comparison between the FC Layer and the Output Layer

Characteristics	FC Layer	Output Layer
Function	Feature integration; dimension transformation	Task-adapted prediction
Typical activation functions	ReLU/Tanh	Softmax/Sigmoid/Linear
Number of parameters	Need optimization	Small
Task dependence	General-purpose	Dependent on task type

Table 2 Comparison of Common Activation Functions

Activation Function	Multi-Class	Feature Description
ReLU (Rectified Linear Unit)	Yes	When the input is greater than 0, the output is the same as the input; when the input is less than or equal to 0, the output is 0. Calculation is simple, and can effectively alleviate gradient vanishing problem, but may cause some neurons to "die"[18] .
Sigmoid	Yes	The output range is (0,1), which can map the input to the (0,1) interval. It is often used in the output layer of binary classification problems, but it has gradient vanishing problem and high computational complexity[19] .
Tanh (Hyperbolic Tangent Function)	Yes	The output range is (-1,1), which is a variant of Sigmoid. The output range is more symmetric than that of Sigmoid, also having gradient vanishing problem[20] .
Softmax	Yes	The output range is (0,1), and the sum of all outputs is 1. It is often used in the output layer of multi-classification problems and can map to a probability distribution[21] .
Leaky ReLU	Yes	It is an improved version of ReLU. When the input is less than or equal to 0, the output is the product of a very small slope and the input, avoiding the "death problem" of ReLU, but has more parameters[22] .
ELU (Exponential Linear Unit)	Yes	When the input is greater than 0, the output is the same as the input; when the input is less than or equal to 0, the output is a(e^x-1), where a is a hyperparameter, which can alleviate the gradient vanishing problem, but the computational complexity is relatively high[23] .
SELU (Scaled Exponential Linear Unit)	Yes	An improved version of ELU, with self-normalization characteristics (can keep mean and variance stable) and relatively high computational complexity[24] .
PReLU (Parametric ReLU)	Yes	An improved version of Leaky ReLU. Slope parameter can be learned, and the flexibility is higher, but more parameters[25] .

Table 3 Summary of CNN Structure

Part Name	Function	Main Characteristics
Input Layer	Receive input data	Image data (height × width × number of channels)

Convolutional Layer	Extract local features	Use convolutional kernels to slide and output feature maps
Pooling Layer	Down-sample and reduce dimension	Max pooling or average pooling
Activation Layer	Introduce non-linearity	Apply activation functions element-by-element
FC Layer	Flatten feature maps for classification or regression	Neurons are fully connected to the previous layer
Output Layer	Output final results	Classification or regression tasks
Loss Function	Measure the difference between model output and true labels	Cross-entropy loss for classification, mean squared error for regression
Optimizer	Update network parameters to minimize the loss function	Update parameters through backpropagation

3. Improvements of CNN

In 1998, LeCun et al. proposed LeNet-5, which established the basic architecture of CNN. This network integrated convolutional layers, pooling layers, and FC layers for the first time and was successfully applied to the handwritten digit recognition task[9]. In 2012, AlexNet proposed by Krizhevsky et al. achieved a breakthrough in the ImageNet competition by introducing ReLU activation functions, Dropout technology, and multi-GPU parallel training, solving the problems of gradient vanishing and overfitting[18]. In 2014, the VGG network developed by Simonyan and Zisserman adopted 3×3 small convolution kernels and a 16-19-layer deep structure, proving the performance improvement brought by depth[26]; in the same year, GoogLeNet proposed by Szegedy et al. achieved parallel parallel extraction of multi multi-scale features through the Inception module, greatly reducing the number of parameters[11]. In 2015, He et al. proposed ResNet, which broke through the depth limitation through a residual learning framework and achieved stable training of a 152-layer deep network for the first time[27]. In 2017, Huang et al. further proposed DenseNet, which used dense connections to enhance feature reuse [17]. In 2018, significant progress was made in lightweight architectures for mobile devices: MobileNetV2 by Sandler et al. introduced depthwise separable convolutions and inverted residual structures[28], while ShuffleNet by Zhang et al. achieved efficient computation through group convolutions and channel shuffling[29]. The Deep Residual Shrinkage Network proposed in 2019 enhanced the noise resistance through soft thresholding operations[30]. In the same year, ESPNetV2, designed specifically for edge devices, significantly reduced the computational complexity by using depthwise separable dilated convolutions[31], and in the lightweight direction, FBNet V1 achieved lightweight model optimization through differentiated neural architecture search[32].

In 2020, Vision Transformer (ViT) successfully applied the self-attention mechanism to visual tasks for the first time, leading the transformation of CNN towards a general-purpose architecture [33]. In 2021, Touvron et al. proposed MoCo v3, a self-supervised learning method, further enhancing performance in unsupervised learning tasks[34]. In 2022, ConvNeXt brought the performance of traditional CNN close to that of ViT through modernization modifications such as 7×7 large convolution kernels and LayerNorm[35]. In 2025, the OverLoCK architecture imitated the human visual mechanism to achieve long-range dependency modeling[36]. Recently, methods for automatically designing neural network architectures have emerged, greatly reducing the impact of manual intervention. For example, in 2018, Liu et al. proposed Neural Architecture Search[37], evaluate performance through training and gradually optimize the architecture design through strategies like reinforcement learning, Bayesian optimization.

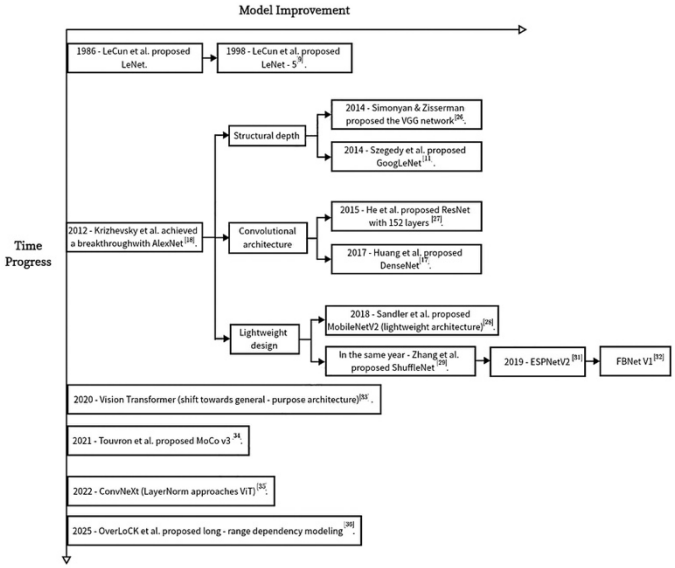

Figure 2: Classic Models of CNN and Its Variants

4. Applications of CNN in Medicine

In recent years, CNNs have made remarkable progress in medical applications and have become important tools in the fields of medical image analysis and auxiliary diagnosis. Ker et al.'s 2017 research systematically reviewed the early applications of artificial intelligence in medical imaging[38]. In 2019, the research team of Han and Deepak et al. supplemented relevant research results from different perspectives[39, 40] respectively. Building on previous studies, this paper focuses on sorting out the latest research progress from 2019 to 2025, mainly concentrating on classification, boundary determination, registration, and natural language text processing.

4.1 Classification

In medical image classification tasks, CNN, with its powerful feature learning ability, can effectively distinguish between benign and malignant lesions and achieve automated identification of multiple diseases. This technology establishes a reliable classification model by extracting multi-level features of images, providing important auxiliary decision-making support for clinical diagnosis. In 2020, Kandel et al. systematically studied the impact of hierarchical fine-tuning strategies in transfer learning on classification performance. Through verification with a histopathology dataset, they found that moderately fine-tuning the top-level network blocks can not only maintain the robustness of the model but also significantly reduce the computational cost[41]. Aiming at the clinical difficulty of brain tumors, in 2021, Irmak innovatively constructed a three-level CNN classification system. By automatically optimizing hyperparameters through a grid search algorithm, it achieved an accuracy rate of 92.66% in the five-type tumor classification task, and its performance surpassed traditional architectures such as ResNet[42]. In the same year, the Das team introduced a genetic algorithm into diabetic retinopathy classification, realizing the automated design of key parameters such as the number of CNN layers and the size of convolution kernels, and achieved an AUC of 0.9933 on the Messidor dataset[43].

With the development of technology, new network architectures continue to emerge. In 2022, the GATE-CNN model proposed by Mishra innovatively integrated the graph attention mechanism. On three brain tumor MRI datasets, the average accuracy rate reached 98.96%, significantly outperforming baseline models such as traditional multi-input CNN[44]. For time-series signal classification, in 2022, Huang used fuzzy integration to fuse the decision results of multiple CNN models and optimized the confidence weights through particle swarm optimization, effectively improving the single-trial recognition accuracy of motor imagery EEG signals[45]. The combination of intelligent optimization algorithms and new network architectures has enabled CNN to

continuously break through performance bottlenecks in medical classification tasks. On this basis, in 2023, Ghosh et al. further optimized the automated network design method and proposed the TPEvo-CNN model, which innovatively combined differential evolution and genetic algorithms for CNN architecture search[46]. At the same time, to the key clinical requirement of model interpretability, Aswath et al. designed a three-branch self-explanatory TS-CNN. Through the collaborative learning of global-local-fusion branches, achieving a dual improvement in performance and interpretability in skin lesion classification task[47].

The latest research trends show that model efficiency and multi-modal feature fusion have become new breakthrough directions. In 2024, the Eff-CTM network proposed by the Liu team groundbreakingly integrated the advantages of CNN, Transformer, and MLP. Through a multi-stage feature extraction strategy, while maintaining lightness, it significantly improved the classification accuracy of diseases such as pneumonia and colon cancer[48]. It is worth noting that in 2025, Amoury applied particle swarm optimization to brain tumor MRI classification. The PSO-CNN model obtained through automated architecture search achieved a classification accuracy rate of 99.19%, pointing out the direction for the development of lightweight and high-precision diagnostic systems[49].

Table 4: Datasets Used for Classification

Dataset	Year	Link	Publish	Use
BCIC-IV-2a	2008	Unavailable	[64]	[45]
ImageNet	2010	Unavailable	[62]	[41]
Messidor	2014	Unavailable	[63]	[43]
PatchCamelyon Histopathology	2015	http://www.bioimaging2015.ineb.up.pt/dataset.html	-	[41]
RIDER	2015	https://doi.org/10.7937/K9/TCIA.2015.VOSN3HN1	-	[42]
REMBRANDT	2015	https://doi.org/10.7937/K9/TCIA.2015.588OZUZB	-	[42]
Brain 1	2015	https://doi.org/10.1371/journal.pone.0140381	-	[42]
TCGA-LGG	2016	https://doi.org/10.7937/K9/TCIA.2016.L4LTD3TK	-	[42]
Chest-Xray	2018	https://data.mendeley.com/datasets/rscbjbr9sj/2	-	[48]
Small-ISIC2018	2018	https://www.cvmart.net/dataSets/detail/729?channel_id=op10	[70]	[48]
COVID-19 1	2020	Unavailable	[65]	[46]
COVID-19 2	2020	https://data.mendeley.com/datasets/8h65ywd2jr/3	-	[46]
COVID-19 3	2020	Unavailable	[66]	[46]
COVID-19 4	2020	Unavailable	[67]	[46]
Pneumonia	2020	Unavailable	[68]	[46]
PAD-UFES-20	2020	Unavailable	[69]	[47]
Skin cancer	2021	https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign	-	[46]
Chaoyang	2021	https://bupt-ai-cz.github.io/HSA-NRL/	[71]	[48]
MRI 1	2021	https://www.kaggle.com/saadalbawi/brain-tumor-mri-dataset	-	[44]
MRI 2	2021	https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset	-	[49]

4.2 Boundary Detection

In medical image analysis, precise boundary detection is of crucial significance for clinical diagnosis and treatment. Especially in surgical planning such as tumor resection, accurate delineation of lesion boundaries directly affects the formulation of surgical plans and treatment outcomes. In 2022, Shi et al. proposed a hybrid method combining multi-channel CNN and fuzzy active contour model, achieving good robustness in lung tumor segmentation[50]. In the same year, to further improve segmentation performance, Chen's team developed Posterior-CRF, combining features learned by CNN with the traditional CRF framework[51]. Aiming at the limitations of U-Net, Wang et al. proposed the O-Net hybrid architecture by integrating the advantages of Swin Transformer and CNN,

making breakthrough on the Synapse multi-organ CT dataset[52]. With the development of weakly supervised learning, the ScribFormer model proposed by Li's team in 2024 innovatively adopted a three-branch structure, effectively solving the problem of global shape information learning under scribble annotations[53]. At the same time, HCT-Unet developed by Fan et al. achieved a Dice coefficient of 82.23% in multi-organ segmentation tasks and even set a new record of 91% in heart segmentation[54]. These studies jointly promote the development of medical image segmentation technology.

Table 5: Datasets Used for Boundary Detection

Dataset	Year	Link	Publish	Use
CT Arteries	2009	https://www.sciencedirect.com/science/article/pii/S1556086415316786	[72]	[51]
ISLES	2017	http://refhub.elsevier.com/S1361-8415(21)00356-X/sbref0020	[73]	[51]
ACDC	2018	Unavailable	[77]	[53]
ISIC2017 skin lesion challenge	2018	Unavailable	[75]	[52]
MSCMRseg	2019	Unavailable	[78]	[53]
WMH	2019	http://refhub.elsevier.com/S1361-8415(21)00356-X/sbref0014	[74]	[51]
Synapse multi-organ segmentation	2021	Unavailable	[76]	[52]

4.3 Registration

Medical image registration is of great significance in clinical diagnosis and treatment. Especially in scenarios such as neurosurgical navigation and spinal surgery planning, precise alignment of anatomical structures directly affects the surgical outcome. Deep learning-based registration methods can directly learn spatial transformation parameters from image data, significantly improving the registration efficiency and accuracy, and also demonstrating adaptability to multi-modal images.

In 2022, Song et al. propose a hybrid registration network combining Transformer and CNN by integrating global and local information through an encoder-decoder architecture[55]. On this basis, in 2023, Zhu's team introduced a similarity attention mechanism and developed a 3D registration method based on multi-scale CNN, demonstrating excellent robustness in challenging tasks such as large-range registration of abdominal CT[56]. In the same year, the Information Complementary Network (ICN) proposed by Deng further improved the Dice coefficient by 3% in 3D brain MRI registration through grid block embedding and dual-path decoder design. At the same time, the number of folding points was reduced by 50%, significantly improving the smoothness of the deformation field[57]. The demand for real-time performance has promoted the development of hardware acceleration technologies. Subsequently, aiming at the real-time registration needs in ultrasound-guided surgery, Aydin's team developed the FUIR-CNN system based on FPGA acceleration, providing a feasible solution for real-time intraoperative registration[58].

Table 6: Datasets Used for Registration

Dataset	Year	Link	Publish	Use
Brain 2	2005	Unavailable	[80]	[56]
OASIS	2007	Unavailable	[79]	[55, 57]
Brain 3	2014	Unavailable	[81]	[56]
Brain 4	2017	Unavailable	[82]	[56]
Abdominal	2019	https://arxiv.org/abs/1902.09063	-	[56]

Chest	2020	https://www.cancerimagingarchive.net/collection/lidc-idri/	-	[56]
LiTS	2023	Unavailable	[83]	[57]
Breast Ultrasound Lesions	2017	Unavailable	[84]	[58]
BUSI	2020	Unavailable	[85]	[58]
B-Mode Common Carotid Artery Ultrasound Image	2023	http://splab.cz/en/download/databaze/ultrasound	-	[58]

4.4 Natural Language Text Processing

Traditional medical report analysis relies highly on manual reading and annotation, which is inefficient and subject to subjective bias. In recent years, CNN combined with natural language processing technology has shown unique advantages in clinical text analysis: Through its hierarchical feature extraction capability, CNN can effectively capture key diagnostic information in medical texts, including structured content such as disease characteristics, examination results, and treatment plans. This automated label extraction technology not only greatly improves the efficiency of clinical data annotation but also provides a new paradigm for building high-quality medical image-text and related datasets. In 2019, Yao et al. pioneered the combination of UMLS medical entity embedding with CNN and verified the effectiveness of the knowledge-guided CNN model in the i2b2 challenge[59]. Aiming at the complexity of Chinese medical question answering, in 2020, the Zhang team proposed a BiGRU-CNN hybrid architecture. By fusing the local feature extraction of CNN and the sequence modeling advantage of GRU, it achieved a breakthrough improvement in question-answering matching accuracy on the cMedQA dataset[60].

With the in-depth application of electronic health records (eHR), in 2023, the Yang team developed two innovative models, AT-CNN and IT-CNN, providing a feasible solution for the real-time processing of clinical texts [61]. These studies collectively demonstrate the diverse applications of CNN in medical text processing.

References

[1] Erickson, B. J., Korfiatis, P., Akkus, Z., & Kline, T. L. (2017). Machine learning for medical imaging. radiographics, 37(2), 505-515.

[2] Sarvamangala, D. R., & Kulkarni, R. V. (2022). Convolutional neural networks in medical image understanding: a survey. Evolutionary intelligence, 15(1), 1-22.

[3] Zarándy, Á., Rekeczky, C., Szolgay, P., & Chua, L. O. (2015, May). Overview of CNN research: 25 years history and the current trends. In 2015 IEEE international symposium on circuits and systems (ISCAS) (pp. 401-404). IEEE.

[4] Yadav, S. S., & Jadhav, S. M. (2019). Deep convolutional neural network based medical image classification for disease diagnosis. Journal of Big data, 6(1), 1-18.

[5] Salehi, A. W., Khan, S., Gupta, G., Alabduallah, B. I., Almjally, A., Alsolai, H., ... & Mellit, A. (2023). A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope. Sustainability, 15(7), 5930.

[6] Kayalibay, B., Jensen, G., & van der Smagt, P. (2017). CNN-based segmentation of medical imaging data. arXiv preprint arXiv:1701.03056.

[7] Boveiri, H. R., Khayami, R., Javidan, R., & Mehdizadeh, A. (2020). Medical image registration using deep neural networks: a comprehensive review. Computers & Electrical Engineering, 87, 106767.

[8] McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, 5, 115-133.

[9] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (2002). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

[10] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

[11] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

[12] Liang, M., Zhou, T., Zhang, F., Yang, J., & Xia, Y. (2018). Convolutional neural network and its application in medical image analysis. Journal of Biomedical Engineering, 35(6), 977.

[13] Martín, A., Ashish, A., Paul, B., Eugene, B., Zhifeng, C., Craig, C., ... & Martin, W. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow. org, 7.

[14] Cheng, X., Wang, P., & Yang, R. (2019). Learning depth with convolutional spatial propagation network. IEEE transactions on pattern analysis and machine intelligence, 42(10), 2361-2379.

[15] Ji, S., Xu, W., Yang, M., & Yu, K. (2012). 3D convolutional neural networks for human action recognition. IEEE transactions on pattern analysis and machine intelligence, 35(1), 221-231.

[16] Nirthika, R., Manivannan, S., Ramanan, A., & Wang, R. (2022). Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study. Neural Computing and Applications, 34(7), 5321-5347.

[17] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

[18] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

[19] Ying, Y., Zhang, N., Shan, P., Miao, L., Sun, P., & Peng, S. (2021). PSigmoid: Improving squeeze-and-excitation block with parametric sigmoid. Applied Intelligence, 51, 7427-7439.

[20] Abdelouahab, K., Pelcat, M., & Berry, F. (2017, September). Why TanH is a hardware friendly activation function for CNNs. In Proceedings of the 11th international conference on distributed smart cameras (pp. 199-201).

[21] Liang, X., Wang, X., Lei, Z., Liao, S., & Li, S. Z. (2017, October). Soft-margin softmax for deep classification. In International Conference on Neural Information Processing (pp. 413-421). Cham: Springer International Publishing.

[22] Xu, J., Li, Z., Du, B., Zhang, M., & Liu, J. (2020, July). Reluplex made more practical: Leaky ReLU. In 2020 IEEE Symposium on Computers and communications (ISCC) (pp. 1-7). IEEE.

[23] Clevert, D. A., Unterthiner, T., & Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (elus). arXiv preprint arXiv:1511.07289.

[24] Zhang, G., & Li, H. (2018). Effectiveness of scaled exponentially-regularized linear units (SERLUs). arXiv preprint arXiv:1807.10117.

[25] Jiang, T., & Cheng, J. (2019, August). Target recognition based on CNN with LeakyReLU and PReLU activation functions. In 2019 international conference on sensing, diagnostics, prognostics, and control (SDPC) (pp. 718-722). IEEE.

[26] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[27] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[28] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

[29] Zhang, X., Zhou, X., Lin, M., & Sun, J. (2018). Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 6848-6856).

[30] Zhao, M., Zhong, S., Fu, X., Tang, B., & Pecht, M. (2019). Deep residual shrinkage networks for fault diagnosis. IEEE Transactions on Industrial Informatics, 16(7), 4681-4690.

[31] Mehta, S., Rastegari, M., Shapiro, L., & Hajishirzi, H. (2019). Espnetv2: A light-weight, power efficient, and general purpose convolutional neural network. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 9190-9200).

[32] Wu, B., Dai, X., Zhang, P., Wang, Y., Sun, F., Wu, Y., ... & Keutzer, K. (2019). Fbnet: Hardware-aware efficient convnet design via differentiable neural architecture search. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10734-10742).

[33] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

[34] Chen, X., Xie, S., & He, K. (2021). An empirical study of training self-supervised vision transformers. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 9640-9649).

[35] Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A convnet for the 2020s. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11976-11986).

[36] Lou, M., & Yu, Y. (2025). OverLoCK: An Overview-first-Look-Closely-next ConvNet with Context-Mixing Dynamic Kernels. arXiv preprint arXiv:2502.20087.

[37] Liu, C., Zoph, B., Neumann, M., Shlens, J., Hua, W., Li, L. J., ... & Murphy, K. (2018). Progressive neural architecture search. In Proceedings of the European conference on computer vision (ECCV) (pp. 19-34).

[38] Ker, J., Wang, L., Rao, J., & Lim, T. (2017). Deep learning applications in medical image analysis. Ieee Access, 6, 9375-9389.

[39] Han, D., Li, Q. H., Cai, W., Xia, Y. W., Ning, J., & Huang, F. (2019). Research and application of artificial intelligence in medical imaging. Big Data Research, 5(1), 2019004.

[40] Deepak, S., & Ameer, P. M. (2019). Brain tumor classification using deep CNN features via transfer learning. Computers in biology and medicine, 111, 103345.

[41] Kandel, I., & Castelli, M. (2020). How deeply to fine-tune a convolutional neural network: a case study using a histopathology dataset. Applied Sciences, 10(10), 3359.

[42] Irmak, E. (2021). Multi-classification of brain tumor MRI images using deep convolutional neural network with fully optimized framework. Iranian Journal of Science and Technology, Transactions of Electrical Engineering, 45(3), 1015-1036.

[43] Das, S., & Saha, S. K. (2022). Diabetic retinopathy detection and classification using CNN tuned by genetic algorithm. Multimedia Tools and Applications, 81(6), 8007-8020.

[44] Mishra, L., & Verma, S. (2022). Graph attention autoencoder inspired CNN based brain tumor classification using MRI. Neurocomputing, 503, 236-247.

[45] Huang, J. X., Hsieh, C. Y., Huang, Y. L., & Wei, C. S. (2022). Toward CNN-based motor-imagery EEG classification with fuzzy fusion. International Journal of Fuzzy Systems, 24(8), 3812-3823.

[46] Ghosh, A., Jana, N. D., Das, S., & Mallipeddi, R. (2023). Two-phase evolutionary convolutional neural network architecture search for medical image classification. IEEE Access, 11, 115280-115305.

[47] Ashwath, V. A., Sikha, O. K., & Benitez, R. (2023). TS-CNN: a three-tier self-interpretable CNN for multi-region medical image classification. IEEE access, 11, 78402-78418.

[48] Liu, S., Wang, L., & Yue, W. (2024). An efficient medical image classification network based on multi-branch CNN, token grouping Transformer and mixer MLP. Applied Soft Computing, 153, 111323.

[49] El Amoury, S., Smili, Y., & Fakhri, Y. (2025). Design of an Optimal Convolutional Neural Network Architecture for MRI Brain Tumor Classification by Exploiting Particle Swarm Optimization. Journal of Imaging, 11(2), 31.

[50] Shi, Q., Yin, S., Wang, K., Teng, L., & Li, H. (2022). Multichannel convolutional neural network-based fuzzy active contour model for medical image segmentation. Evolving Systems, 13(4), 535-549.

[51] Chen, S., Gamechi, Z. S., Dubost, F., van Tulder, G., & de Bruijne, M. (2022). An end-to-end approach to segmentation in medical images with CNN and posterior-CRF. Medical image analysis, 76, 102311.

[52] Wang, T., Lan, J., Han, Z., Hu, Z., Huang, Y., Deng, Y., ... & Chen, G. (2022). O-Net: a novel framework with deep fusion of CNN and transformer for simultaneous segmentation and classification. Frontiers in neuroscience, 16, 876065.

[53] Li, Z., Zheng, Y., Shan, D., Yang, S., Li, Q., Wang, B., ... & Shen, D. (2024). Scribformer: Transformer makes cnn work better for scribble-based medical image segmentation. IEEE Transactions on Medical Imaging, 43(6), 2254-2265.

[54] Fan, Y., Song, J., Yuan, L., & Jia, Y. (2025). HCT-Unet: multi-target medical image segmentation via a hybrid CNN-transformer Unet incorporating multi-axis gated multi-layer perceptron. The Visual Computer, 41(5), 3457-3472.

[55] Song, L., Liu, G., & Ma, M. (2022). TD-Net: unsupervised medical image registration network based on Transformer and CNN. Applied Intelligence, 52(15), 18201-18209.

[56] Zhu, F., Wang, S., Li, D., & Li, Q. (2023). Similarity attention-based CNN for robust 3D medical image registration. Biomedical Signal Processing and Control, 81, 104403.

[57] Deng, L., Zou, Y., Huang, S., Yang, X., & Wang, J. (2023). Deformable 3D medical image registration with convolutional neural network and transformer. Journal of Instrumentation, 18(04), P04029.

[58] Aydin, S. G., & Bilge, H. Ş. (2023). Fpga implementation of image registration using accelerated cnn. Sensors, 23(14), 6590.

[59] Yao, L., Mao, C., & Luo, Y. (2019). Clinical text classification with rule-based features and knowledge-guided convolutional neural networks. BMC medical informatics and decision making, 19, 31-39.

[60] Zhang, Y., Lu, W., Ou, W., Zhang, G., Zhang, X., Cheng, J., & Zhang, W. (2020). Chinese medical question answer selection via hybrid models based on CNN and GRU. Multimedia tools and applications, 79, 14751-14776.

[61] Yang, Z., & Emmert-Streib, F. (2023). Threshold-learned CNN for multi-label text classification of electronic health records. IEEE Access, 11, 93402-93419.

[62] Pan, S.J.; Yang, Q. A Survey on Transfer Learning. IEEE Trans. Knowl. Data Eng. 2010, 22, 1345–1359.

[63] Decencière E, Zhang X, Cazuguel G, Lay B, Cochener B, Trone C, Gain P, Ordonez R, Massin P, Erginay A et al (2014) Feedback on a publicly distributed image database: the messidor database. Image Analysis & Stereology 33(3):231

[64] Brunner, C., Leeb, R., Mu¨ller-Putz, G., Schlo¨gl, A., Pfurtscheller, G.: ‘‘Bci competition 2008-graz data set a,’’ Institute for Knowledge Discovery (Laboratory of Brain-Computer Inter- faces). Graz University of Technology 16, 1–6 (2008)

[65] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, “Automated detection of COVID-19 cases using deep neural networks with X-ray images,” Comput. Biol. Med., vol. 121, Jun. 2020, Art. no. 103792.

[66] E. Soares, P. Angelov, S. Biaso, M. H. Froes, and D. K. Abe, “Sars-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification,” MedRxiv, p. 2020–04, 2020.

[67] M. E. H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. A. Emadi, M. B. I. Reaz, and M. T. Islam, “Can AI help in screening viral and COVID-19 pneumonia? ” IEEE Access, vol. 8, pp. 132665–132676, 2020.

[68] U. Sait, K. V. G. Lal, S. P. Prajapati, R. Bhaumik, T. Kumar, S. Shivakumar, and K. Bhalla, “Curated dataset for COVID-19 posterior-anterior chest radiography images (X-rays),” Mendeley Data, V3, 2021, doi: 10.17632/9xkhgts2s6.3.

[69] A. G. Pacheco, G. R. Lima, A. S. Salomão, B. Krohling, I. P. Biral, G. G. de Angelo, F. C. Alves Jr., J. G. Esgario, A. C. Simora, P. B. Castro, F. B. Rodrigues, P. H. Frasson, R. A. Krohling, H. Knidel, M. C. Santos, R. B. do Espírito Santo, T. L. Macedo, T. R. Canuto, and L. F. de Barros, ‘‘PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones,’’ Data Brief, vol. 32, Oct. 2020, Art. no. 106221.

[70] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M. Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, et al., Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic), 2019.

[71] Chuang Zhu, Wenkai Chen, Ting Peng, Ying Wang, Mulan Jin, Hard sample aware noise robust learning for histopathology image classification, IEEE Trans. Medical Imaging 41 (4) (2022) 881–894.

[72] Pedersen, J.H., Ashraf, H., Dirksen, A., Bach, K., Hansen, H., Toennesen, P., Thorsen, H., Brodersen, J., Skov, B.G., Døssing, M., et al., 2009. The Danish randomized lung cancer CT screening trial-overall design and results of the prevalence round. J. Thoracic Oncol. 4, 608–614.

[73] Maier, O., Menze, B.H., von der Gablentz, J., Häni, L., Heinrich, M.P., Liebrand, M., Winzeck, S., Basit, A., Bentley, P., Chen, L., et al., 2017. Isles 2015-a public eval- uation benchmark for ischemic stroke lesion segmentation from multispectral MRI. Med. Image Anal. 35, 250–269.

[74] Kuijf, H.J., Biesbroek, J.M., de Bresser, J., Heinen, R., Andermatt, S., Bento, M., Berseth, M., Belyaev, M., Cardoso, M.J., Casamitjana, A., et al., 2019. Standard- ized assessment of automatic segmentation of white matter hyperintensities; results of the WMH segmentation challenge. IEEE Trans. Med. Imaging.

[75] Codella, N. C., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., et al. (2018). "Skin lesion analysis toward melanoma detection: a challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic)," in 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)(Washington, DC: IEEE), 168–172.

[76] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., et al. (2021). Transunet: Transformers make strong encoders for medical image segmentation. arXiv [Preprint] arXiv:2102.04306.

[77] O. Bernard et al., "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?" IEEE Trans. Med. Imag., vol. 37, no. 11, pp. 2514–2525, Nov. 2018.

[78] X. Zhuang, "Multivariate mixture model for myocardial segmentation combining multi-source images," IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 12, pp. 2933–2946, Dec. 2019.

[79] Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL (2007) Open access series of imaging studies (oasis): Cross-sectional mri data in young, middle aged, nondemented, and demented older adults. J Cogn Neurosci 19(9):1498–1507

[80] S.G. Mueller, M.W. Weiner, L.J. Thal, R.C. Petersen, C.R. Jack, W. Jagust, J. Q. Trojanowski, A.W. Toga, L. Beckett, Ways toward an early diagnosis in alzheimers disease: the alzheimers disease neuroimaging initiative (ADNI), Alzheimer's Dementia 1 (1) (2005) 55–66.

[81] A.D. Martino, C. Yan, Q. Li, E. Denio, F.X. Castellanos, K. Alaerts, J.S. Anderson, M. Assaf, S.Y. Bookheimer, M. Dapretto, et al., The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism, Mol. Psychiatry 19 (6) (2014) 659–667.

[82] P. Bellec, C. Chu, F. Chouinard-Decorte, Y. Benhajali, D.S. Margulies, R.C. Craddock, The neuro bureau ADHD-200 preprocessed repository, Neuroimage 144 (B) (2017) 275–286.

[83] P. Bilic, P. Christ, H. Bran Li, E. Vorontsov, A. Ben-Cohen, G. Kaissis et al., The Liver Tumor Segmentation Benchmark (LiTS), Med. Image Anal. 84 (2023) 102680 [arXiv:1901.04056].

[84] Yap, M.H.; Pons, G.; Marti, J.; Ganau, S.; Sentis, M.; Zwiggelaar, R.; Davison, A.K.; Marti, R. Automated Breast Ultrasound Lesions Detection using Convolutional Neural Networks. IEEE J. Biomed. Health Inform. 2017, 22, 1218–1226.

[85] Al-Dhabyani, W.; Gomaa, M.; Khaled, H.; Fahmy, A. Dataset of breast ultrasound images. Data Brief 2020, 28, 104863.