Cancer Detection Using Improved CNN-Based Models

Bo Ning[a]
Hangzhou Dianzi University ITMO Joint Institute, Hangzhou Dianzi University, Hangzhou, 310018, China

Keywords: Convolution Neural Network, Cancer Image Analysis, Transfer Learning, Lightweight Models, Multi-Model Fusion.

Abstract: Medical imaging is a crucial part of clinical diagnosis, yet processing its large-volume data is challenging. Convolutional neural networks (CNNs) have shown great potential in medical image analysis. However, the high computational cost and need for large annotated datasets often limit their widespread clinical adoption This paper focuses on CNN-based models for cancer detection. The paper delves into various innovative models applied in breast, lung, and skin cancer detection. These newly proposed models excel in specific aspects, whether in high-precision classification, classification efficiency, or lightweight design. But some models still face issues like poor generalization on rare diseases and high computational requirements. This paper summarizes these issues and identifies current and future research directions, including the development of generalized cancer detection frameworks and the application of transfer learning techniques. Overall, the paper highlights the enormous potential of CNN-based models in medical imaging while pointing out the need for continuous research and development to overcome existing challenges and limitations.

1 INTRODUCTION

Medical imaging, as an important auxiliary means of clinical diagnosis, is a key link to ensure the success of treatment in the process of medical diagnosis and treatment. it plays an important role in life science research. Medical imaging includes X-rays, ultrasound, computed tomography, magnetic resonance imaging, positron emission tomography, etc. These medical image data are huge in volume and difficult to process in clinical diagnosis. Therefore, automatically detecting diseases from medical images has become a key issue in the medical field (Chen, Mat Isa and Liu, 2025).

Medical image analysis requires powerful algorithms that can extract details from high-dimensional and noisy datasets. Convolutional neural networks are precisely the unique options that can address this challenge. Convolutional Neural Network (CNN) is a multi-layer neural network used to extract visual patterns from pixel images. It can automatically extract and select features and classify them, providing radiologists with faster and more accurate diagnostic results in real time. In the field of medical image analysis, CNN has become an advanced algorithm for tasks such as disease detection, organ segmentation and image enhancement (Chen, Mat Isa and Liu, 2025) (Patel and Khan, 2022) (Mienye, Swart and Obaido et.al, 2025). The convolutional neural network technology of deep learning is widely applied in the analysis of medical images. It is particularly used for diagnosing different types of diseases, such as breast cancer, Alzheimer's disease, brain tumors, and so on. Algorithms based on deep convolutional neural networks have achieved remarkable results in the analysis of medical images. For medical image data, various types of transfer learning methods have been proposed and have achieved remarkable results, such as AlexNet, VGGNet, ResNet, GoogleNet, etc (Salehi, Khan and Gupta. et al, 2023).

However, there are still many challenges in the application of CNN in the medical field, which hinders its wide application in clinical practice(Mienye, Swart and Obaido et.al, 2025). Some models still face challenges such as insufficient generalization ability on rare diseases or niche datasets, high demands for computing resources, and difficulties in multimodal data fusion. This clarifies that CNN still needs to continuously innovate in

[a] https://orcid.org/0009-0009-3995-4964

Ning, B.
Cancer Detection Using Improved CNN-Based Models.
DOI: 10.5220/0014323000004718

In *Proceedings of the 2nd International Conference on Engineering Management, Information Technology and Intelligence (EMITI 2025)*, pages 138-144
ISBN: 978-989-758-792-4

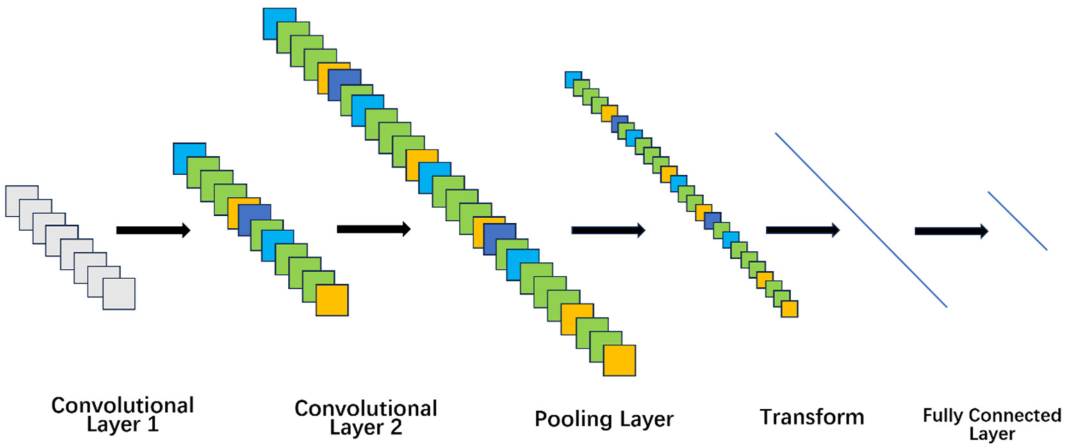

Figure 1: Basic architecture of CNN

aspects such as architectural design, interpretable AI technology, cross-modal and cross-domain generalization, to overcome all these difficulties.

This study will conduct a comprehensive review of the CNN model framework corresponding to the diagnosis of various types of cancers. It will first introduce the working principle of CNN, and then analyze the application cases of CNN in the diagnosis of breast cancer, lung cancer and skin cancer respectively. The paper will focus on elaborating the cutting-edge model architecture based on CNN used therein, as well as the advantages of these models and the achievements obtained. This review aims to explore strategies for improving the performance of cancer diagnosis models, and provide researchers with guidance on applying CNN to solve medical problems such as cancer diagnosis and analysis.

2 THE PRINCIPLE OF CNN

As Figure 1 shows, as a hierarchical and cascade model, the CNN starts by capturing pixel-level information from the lowest layer of the input image matrix and progressively extracts critical feature information through each subsequent layer. Convolution, pooling, and fully connected layers are the three key components of CNN architecture. The convolution layer uses a convolutional operation of filters on the input image to extract local information. After that, higher-level features will be extracted by moving the acquired feature mappings to the subsequent convolution layer. After the convolutional layer, the image dimension is decreased using the pooling layer, also known as the down sampling layer. Because max-pooling lowers the dimension while keeping the image's primary feature, it is frequently employed as a down sampling technique. After the last convolutional layer, the fully connected layer flattens the feature map into a vector. To normalize the output and produce a final output in probability form, a softmax function is often applied at the output layer. This probability indicates the picture belongs to a particular category. CNN typically updates the convolution kernel and fully connected layers' weights during the training phase using a gradient descent approach. The learning method updates the weights until convergence by backpropagating a classification loss into the network (Chen, Mat Isa and Liu, 2025).

3 THE APPLICATION OF CNN IN BREAST CANCER DETECTION

3.1 Swincnn Fusion Transformer Realizes Classification of Multi-Subtype Tumors

V. Sreelekshmi et al. proposed the SwinCNN architecture, integrating Depthwise Separable Convolutional Neural Network and Swin Transformer (Sreelekshmi Pavithran and Nair, 2024). This model realizes the high-precision classification of benign and malignant tumors and their subtypes. The proposed mode SwinCNN, is shown in Figure 2. This architecture consists of two channels: the lower channel is the Local Feature Extraction module (LFM), and the upper channel is the Global Feature Extraction module (GFM). The convolutional layer in

GFM is the main feature extractor, consisting of 64 7 ×7 filters. It is used to generate the feature maps of the input data. The convolution kernels in its three residual blocks are of size 3×3, and the number of filters is 64, 128, and 256, respectively. The final feature map of GFM is downsampled and fed into the Swin Transformer block to extract context information. LFM employs three different types of depthwise separable convolutional blocks. Finally, the outputs of GFM and LFM were combined and input into the softmax layer to calculate the classification probability for different breast cancer types. The result corresponding to the highest probability was taken as the final classification.

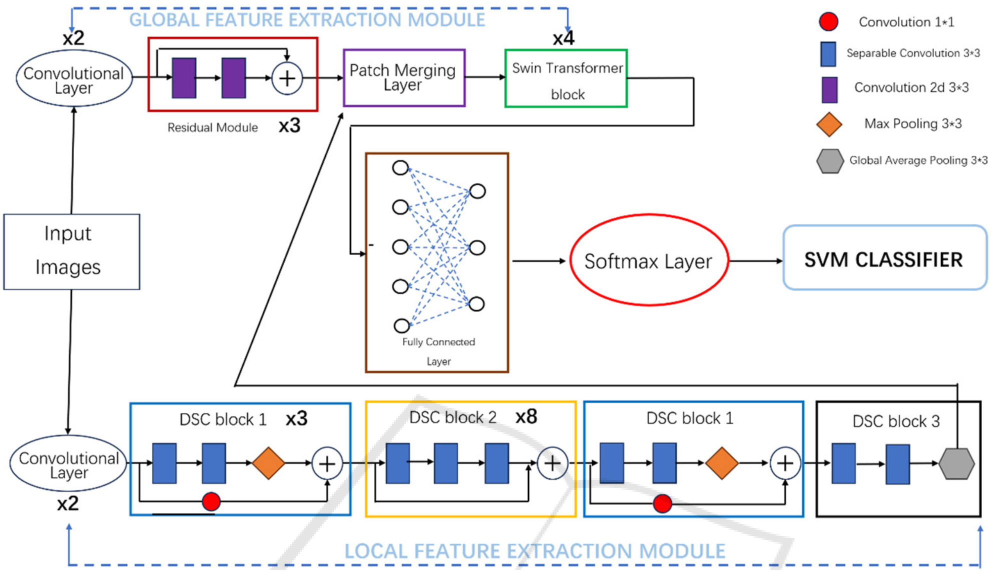

Figure 2: The architecture of SwinCNN

The innovative transformation of CNN is one of the core technical contributions of the proposed model. Traditional CNNs use standard convolutions, which are inefficient in high-resolution scenes of medical images. Their computation scales cubically with the number of channels and the size of the convolution kernel. V. Sreelekshmi et al. proposed a scheme that fuses GoogleNet and Xception models. GoogleNet provides multi-scale features, and Xception ensures computational efficiency. The integration of the two enriches the representation of local features (covering details at different scales) and avoids substantial computational overhead. This makes the model more comprehensive and lightweight on local feature extraction. The traditional CNN is limited to the local receptive field, which can only capture the local cell morphology and fail to model the global structure of the tumor tissue. This leads to difficulties in differentiating invasive carcinomas from carcinomas in situ. SwinCNN compensates for this shortcoming with the Transformer path. On the BACH dataset, the recall of invasive cancer classification is improved from 85% of ResNet to 91.4%.

Finally, the model proposed by V. Sreelekshmi et al. was validated on three major public datasets. Experimental results show that the model achieves extremely high accuracy in detecting various tumors and their subtypes, and demonstrates strong generalization ability for clinically critical subtypes (Sreelekshmi Pavithran and Nair, 2024).

3.2 UWB-CNN-LSTM Enables Early Tumor Detection and Localization

Min Lu et al. conducted research on the early detection and localization of breast cancer. They developed an end-to-end framework based on ultra-wideband (UWB) microwave technology and deep learning to achieve automatic detection of breast cancer and breast quadrant localization. The CNN-LSTM hybrid network framework proposed has achieved full-process automated feature learning. LSTM processes the time domain signals from three-channel signals, and remembers the time dependence of tumor responses through a gating mechanism. This effectively alleviates the problem of vanishing gradients in traditional RNNs. CNN automatically captures the local waveform features of UWB signals through two layers of convolution, replacing the traditional manual extraction of time-domain and frequency-domain features (Lu, Xiao and Pang et al,2022).

The scheme proposed by Min Lu et al. mainly has three advantages. The model adopts a lightweight network architecture, using a shallow CNN in combination with an LSTM. This strikes a balance between computational efficiency and accuracy. The model's training time on a 16-core CPU is only 1012 seconds, making it suitable for clinical scenarios with limited resources. The model also reduces the cost of breast cancer screening. The UWB devices adopted in the scheme have a much lower cost than MRI, making them suitable for popularization in grassroots healthcare settings. The proposed model can also achieve full-process automation from signal input to quadrant positioning, reducing human interpretation errors.

Finally, through multi-scenario verification, the proposed model achieves high-accuracy tumor detection (99.56%) and quadrant localization (F1 score $\geq$ 97%). The CNN-LSTM hybrid network framework enables tumor detection and breast quadrant localization while mitigating issues of high costs, radiation hazards, and tedious manual feature engineering in traditional methods. The model requires a large amount of data for constructing the dataset of the training network. This is a problem that hinders the application of the proposed method in clinical practice.

3.3 Binary Classification for Breast Cancer Based on Multi-Model Integration

Samriddha Majumdar et al. Integrated three models, namely GoogleNet, VGG-11 and MobileNetV3, to solve the problem of binary classification (benign and malignant) of breast cancer pathological images. GoogleNet uses Inception blocks for multi-scale feature extraction (Majumdar, Pramanik and Sarkar, 2023). Large convolution kernels capture the overall information of the image, while small convolution kernels focus on capturing fine details. This enables efficient fusion of multi-scale features. VGG11's shallow architecture captures local details, and its low-depth design reduces the number of parameters of the model. MobileNetV3_Small's bottleneck layers enable efficient high-magnification analysis.

Aiming at the problems of insufficient generalization ability of a single model and fixed weight of a traditional ensemble model, they adopted a fusion strategy based on the Gamma function to achieve more accurate classification. The proposed scheme uses transfer learning to make up for the shortcomings of CNN which requires massive data and long time to train medical images. The test results of the model are also remarkable. The model performs well on BreakHis (four magnifications) and ICIAR-2018 datasets. This proves that the method does not rely on a specific data distribution, and has excellent generalization ability and robustness.

4 APPLICATION OF CNN IN LUNG CANCER DETECTION

4.1 Lightweight CNN Achieves Efficient Classification of Lung Cancer

Mohd Mohsin Ali et al. focus on the application of lightweight deep learning models in medical imaging. They developed a lightweight and efficient CNN model to enable automatic classification of lung cancer (benign, malignant, and normal), mitigating the issues of high computational costs and difficult edge device deployment in traditional models (Ali, Jain and Chauhan et al,2023).

Figure 3 shows the framework of the proposed model. The first convolutional layer (Conv2D) has 32 6×6 kernels with a step size of 3×3 and an activation function of ReLU. This convolution layer extracts basic visual features (e.g., lung nodule edges, texture), and the output feature map size is 169×169×32. The ReLU activation function avoids gradient disappearance while introducing nonlinearity and has low computational cost, which is suitable for lightweight design. The second convolutional layer (Conv2D) also has 32 6×6 kernels with a step size of 3×3. This is to further extract complex texture features, and the output feature map size is 27×27×32. The two-layer convolution balances detail and semantic information through feature extraction at different depths. The flatten layer converts 3D feature maps into a 1D vector (5,408 dimensions), feeding directly into the fully connected layer. The hidden layer contains 128 neurons, reducing the number of parameters through sparse connections. The output layer has 3 neurons (corresponding to three classes: benign, malignant, and normal), with the SoftMax activation function generating class probability distributions.

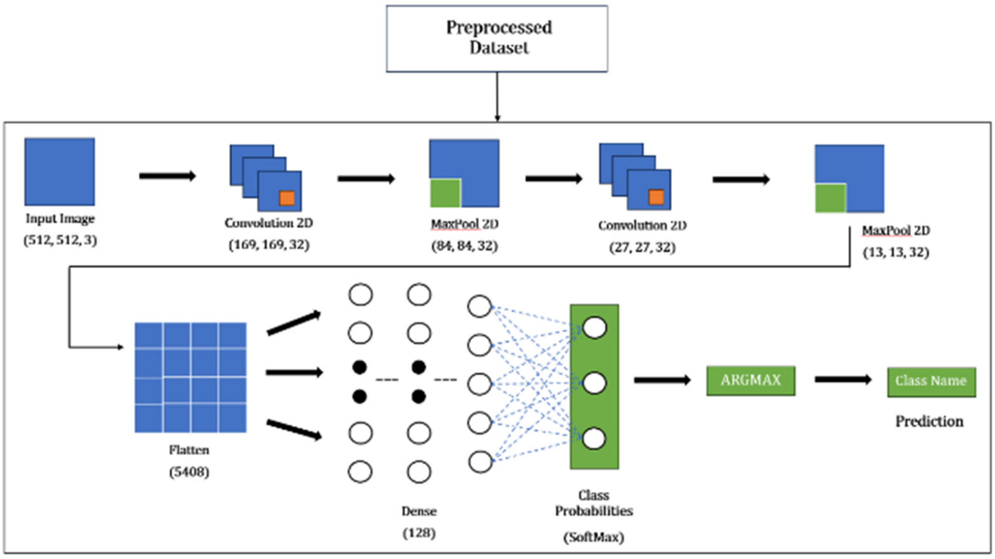

Figure 3: Neural network design diagram.

The model proposed by Mohd Mohsin Ali et al. has several advantages. It uses a lightweight architecture, and the storage size is compressed to 8.43MB through shallow convolution (2 layers), large-step pooling, and low-dimensional fully connected layers. The model can achieve efficient feature extraction, and it uses 6×6 convolution kernel to balance the receptive field and the amount of computation. Finally, the model achieves a validation accuracy of 99%, a training time of 1 minute, and a model size of 8.43MB. Compared with mainstream models such as XceptionNet and MobileNet, it significantly outperforms in all three core metrics(Ali, Jain and Chauhan et al, 2023).

The core contribution of the proposed model lies in breaking through the trade-off between performance and complexity of traditional deep - learning models. While achieving a lightweight design, the model maintains a high accuracy of 99%. This achievement not only provides a new solution for lung cancer screening but also points out the development direction of "lightweight and efficient model".

4.2 Hybrid Detection Model of SMA-CNN and Squeeze-Inception V3

Geethu Lakshmi G et al proposed a hybrid model of "SMA-CNN feature extraction and Squeeze Inception V3 classification" for the accuracy and efficiency of CT image detection of lung cancer (Lakshmi and Nagaraj, 2025). The goal of the proposed model is to increase the detection rate of early-stage lung cancer while maintaining computational lightness. SMA-CNN uses the slime mold algorithm to dynamically adjust the weight of CNN's convolution kernel, which enhances the feature capture of low-contrast tumors and replaces the traditional CNN training with fixed parameters. Squeeze-Inception V3 combines the lightweight design of SqueezeNet with the multi-scale feature extraction of Inception V3. The Fire Module of SqueezeNet compresses the number of channels through a 1×1 convolution kernel, and its parameter number is 1/50 of the parameter number of AlexNet. The average pooling layer of SqueezeNet replaces the fully connected layer, and the decomposition convolution of Inception V3 is combined to further reduce the computational complexity. Inception V3 captures nodule contours and details of different scales by using convolution kernels of different sizes such as 1×1 and 3×3 in parallel. Traditional single-scale convolutions tend to miss detecting small tumors, while the multi-branch structure of Inception V3 enables feature complementation through different receptive fields, achieving a 12% improvement in the recognition rate of small-sized lesions. Finally, the proposed model achieved maximum specificity, accuracy, and sensitivity on the Chest CT-Scan dataset (1,001 cases, covering four classes: adenocarcinoma, large cell carcinoma, squamous cell carcinoma, and normal), with the respective rates of 94%, 95%, and 98%. This model reduces the missed diagnosis and misdiagnosis rates for lung cancer, while the lightweight design enhances its adaptability to resource-constrained environments.

5 APPLICATION OF CNN IN SKIN CANCER DETECTION

Ashwani Kumar realize high precision of skin cancer detection by improving the Falcon finch depth of CNN. The core of the proposed model lies in the combination of ResNet feature transfer and the hybrid optimization algorithm, breaking through the performance bottleneck of traditional CNNS in small samples and complex scenes (Kumar, 2024). The proposed model utilizes Resnet-101 to extract deep features, combines statistical features for dimension reduction processing, forms a 2048-dimensional feature vector, and retains the subtle structural differences in the high-dimensional space. Subsequently, the features are fed into the improved CNN. The Falcon Finch algorithm dynamically adjusts the weights of the fully connected layer through the echolocation mechanism. The FFO algorithm is used to adjust the hyperparameters of the deep CNN classifier, and the optimal combination is determined through 100 iterations. The algorithm optimizes the hyperparameters to improve the efficiency and performance of the classifier, and then improves the accuracy and speed of skin cancer detection. Moreover, the FFO algorithm enhances the robustness of the classifier and accelerates the convergence speed, so that the model can complete the training in a shorter time and achieve better detection performance.

Finally, the experimental results of Ashwani Kumar et al. show that the model optimized by FFO performs well in terms of accuracy, sensitivity and specificity. Ashwani Kumar et al. presented two-index validation results: in k-fold cross validation (k=8), the accuracy, sensitivity, and specificity of the proposed model are 93.59%, 92.14%, and 95.22%, which proves the robustness of the model in small sample scenarios. In the training percentage test (80% data training), the accuracy, sensitivity, and specificity of the proposed model are 96.52%, 96.69%, and 96.54%, which verifies the efficiency under large-scale data. In the comparison experiment, compared with the traditional CNN (accuracy 80.78%), HHO-CNN (86.36%) and SSA-CNN (86.88%), the accuracy of FFO-CNN was increased by 12.81%, 7.23%, and 6.71%, respectively. Its advantage in specificity (distinguishing benign tumors) is even more significant. It proves the effectiveness of FFO in improving the performance of the model. The introduction of Falcon Finch optimization provides a new solution to the problem of parameter tuning of deep neural networks.

The proposed model also faces difficulties. Due to the complex lesion structure, the similar appearance of benign and malignant lesions can lead to difficulties in visual analysis. In the future, hybrid classifiers can be used for skin cancer detection and classification to provide a more comprehensive pathological classification solution.

6 CONCLUSIONS

Convolutional neural networks have promoted the progress of cancer detection, tumor type discrimination, and so on, and show that they still has great potential for development in the medical field. This study further analyzes various CNN-based models for cancer detection image classification and shows the results achieved by each model in each case. This paper presents several solutions for researchers aiming to use CNN models to address cancer detection challenges, helping them understand the corresponding models suitable for various types of cancer detection. Some of these CNN-based models bring higher cancer detection accuracy, some realize the full automation of the detection process, and some have lightweight architectures. While these new CNN-based models have achieved such successes, there are also many problems in their development path: the demand for computing resources of some models is too high, the training of some models still requires a large numbers of data sets to improve the accuracy. And some models can only classify a limited number of categories, resulting in incomplete pathological classification among other issues.

To address these challenges, future research can focus on how to improve network architecture to achieve high accuracy while maintaining model lightweighting, so that the model can adapt to resource-limited grassroots scenarios. Future research should also attempt to explore the transfer learning strategy from skin cancer and breast cancer models to other cancers, and establish a generalized cancer detection framework, so as to make up for the shortcomings of CNN in training medical images that require massive data and extensive training time. to improve the comprehensive judgment capacity of difficult cases, future research might also try to design a multimodal fusion architecture that integrates CT, MRI, pathological pictures, and clinical data to create a full-dimensional cancer diagnostic model. Through improvements in these aspects, CNN will continue to provide better solutions for problems in the medical field.

REFERENCES

Ali, M. M., & Jain, V., & Chauhan, A., & Ranjan, V., & Raj, M. (2023). Automated Lung Cancer Detection Using Lightweight Neural Network. 2023 International Conference on Modeling, Simulation & Intelligent Computing (MoSICom), 219-222.

Chen, C., & Mat Isa, N. A., & Liu, X. (2025). A Review of Convolutional Neural Network Based Methods for Medical Image Classification. Computers in Biology and Medicine, 185, 109507.

Kumar, A., & Kumar, M., & Bhardwaj, V. P., & Kumar, S., & Selvarajan, S. (2024). A Novel Skin Cancer Detection Model Using Modified Finch Deep CNN Classifier Model. Scientific Reports, 14(1), 11235.

Lakshmi G, G., & Nagaraj, P. (2025). Squeeze-Inception V3 with Slime Mould Algorithm-Based CNN Features for Lung Cancer Detection. Biomedical Signal Processing and Control, 100, 106924.

Lu, M., & Xiao, X., & Pang, Y., & Liu, G., & Lu, H. (2022). Detection and Localization of Breast Cancer Using UWB Microwave Technology and CNN-LSTM Framework. IEEE Transactions on Microwave Theory and Techniques, 70(11), 5085-5094.

Majumdar, S., & Pramanik, P., & Sarkar, R. (2023). Gamma Function Based Ensemble of CNN Models for Breast Cancer Detection in Histopathology Images. Expert Systems with Applications, 213, 119022.

Mienye, I. D., & Swart, T. G., & Obaido, G., & Jordan, M., & Ilono, P. (2025). Deep Convolutional Neural Networks in Medical Image Analysis: A Review. Information, 16(3), 195.

Patel, S., & Khan, N. R. (2022). COVID-19 Detection by Medical Images with Pretrained Transfer Learning-Based Model Using CNN: A Systematic Review. Proceedings of 2022 IEEE International Conference on Current Development in Engineering and Technology, CCET 2022.

Salehi, A. W., & Khan, S., & Gupta, G., & Alabduallah, B. I., & Almjally, A., & Alsolai, H., & Siddiqui, T., & Mellit, A. (2023). A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope. Sustainability, 15(7), 5930.

Sreelekshmi, V., & Pavithran, K., & Nair, J. J. (2024). SwinCNN: An Integrated Swin Transformer and CNN for Improved Breast Cancer Grade Classification. IEEE Access, 12, 68697-68710.