

Recent Trends and Advancements in Convolutional Neural Networks for Medical Image Analysis: A Review

Soni¹

Department of Computer Science and Engineering
Guru Jambheshwar University of Sc. &
Technology Hisar, India

Narender Kumar²

Department of Computer Science and Engineering
Guru Jambheshwar University of Sc. &
Technology Hisar, India

Abhishek Kajal³

Department of Computer Science and Engineering
Guru Jambheshwar University of Sc. &
Technology Hisar, India

Abstract - Imaging techniques in medical industry has become one of the main components of this generation's healthcare industry. In the fast few years, Convolutional Neural Network (CNN) has become an important tool and vital piece in the area of medical imaging industry. CNN has revolutionized the way we see the medical pictures and how we analyze it. In this paper we have analyzed the role of CNNs in medical imaging industry. This paper covers and gives insight on the current development and the future of CNN in medical industry. It also features disease-specific applications that show how CNNs may be used to enhance diagnosis and treatment. The review article specifically focuses on the architectures that best suited for the images of diverse modality and inherent complexity.

Keywords - Convolutional Neural Network; Medical Image; Image Processing; deep learning.

I. INTRODUCTION

Medical imaging industry has become completely different all thanks to the CNNs specially the part of automated analysis and diagnosis [1-2]. The main effect has seen in medical imaging of all kinds, including X-rays, MRIs, & CT scans. Medical imaging industry might potentially profit from use of such deep learning (DL) models. With the help of CNNs we can detect anomalies, classify diseases, and assist healthcare professionals by reducing a large spectrum to a more accurate diagnoses [3]. As machines have an ability to learn an adapt from a large data, helps it in improving the efficiency and reliability of medical image interpretation, ultimately enhancing patient care and outcomes. In other words, convolutional Neural Networks (CNNs) have emerged as a groundbreaking innovation in the realm of medical imaging, offering unprecedented capabilities in image analysis and interpretation. These deep learning models have changes the way medical industry work and helps the professionals diagnose and treat patients. By leveraging intricate layers of artificial neurons and specialized convolutional operations [4], [5], CNNs excel at identifying intricate patterns, structures, & anomalies within medical images, including X-rays, & more. With their ability to learn from large datasets, CNNs enhance

the precision and efficiency of image interpretation, enabling earlier disease detection [6], accurate classification, and even the prediction of treatment outcomes [7], [8]. Due to this, CNNs have become one of the most important tools for healthcare providers [9], [10], significantly improvement in patient care and fulfilling the main purpose of saving lives by giving faster, reliable diagnoses and treatment recommendations [11]. The fusion of cutting-edge technology with medical imaging promises to continually redefine the landscape of healthcare [12], [13], fostering a new era of data-driven, & personalized medicine [14].

CNN are obvious choice when it comes to process an image using deep learning models. Their capacity to automatically learn and identify essential features from complex and high-dimensional images has made them particularly useful in the field of medical imaging [15], [16]. Table 1 shows some key aspects of CNNs in medical imaging as found in literature. Following the table, some other key features are discussed to elaborate the role of CNNs in medical imaging.

A. Image Analysis and Diagnosis:

Medical images contain great source of medical information. To extract this information we require expertise to interpret information accurately. CNNs are made to analyze and interpret these images, hence making it one of the most valuable tools for automating the diagnosis of the medical conditions [15].

B. Feature Extraction:

CNNs helps us to automatically identify patterns, edges and structures in the medical images all thanks to use of convolutional layering. These layers helps to identify relevant features, such as tumors, bone fractures, or blood vessel anomalies, without explicit programming [15].

C. Classification and Segmentation:

CNNs are trained to segregate medical images into different categories, where they identify and outline specific regions of interest. For example, they can differentiate between benign

and malignant tumors in mammograms, segment brain tumors in MRI scans, or delineate organs in CT scans [17].

D. Reducing Human Error:

CNNs helps in automating the analyzing the medical images. Automation helps in reading the human labour and human error this making it more reliable and consistent.

E. Enhancing Workflow:

With the help of CNNs we are able to speed up the process of interpretation of medical images [17]. This allows healthcare professionals to focus more on patient care and treatment decisions rather than spending time on image analysis.

F. Large Datasets and Training:

The success of CNNs in medical imaging often relies on the availability of large, annotated datasets. These datasets are used to train the neural networks to recognize specific medical conditions or structures [18].

G. Challenges:

Challenges arrives in every field, with CNNs use their can be issue with data privacy, interpretability. And their is the need for validation and regulation of AI-driven medical diagnostic tools [18].

H. Prospects:

CNNs in medical imaging has come a long way and continue to evolve with improvements in architectures. The architecture includes 3D CNNs for volumetric data, and the integration of CNNs with other AI techniques like natural language processing for better patient diagnosis [18].

Convolutional Neural Networks have brought transformative changes to medical imaging by automating and enhancing the analysis of complex medical images [19], [20]. They are rapidly becoming more and more valuable to the medical community as they improve the speed and precision of diagnosis & treatment planning across a wide range of fields. CNNs excel at applications that need organized grid data, such those seen in photos and videos. They consist of convolutional layers that automatically detect and learn patterns and features within input data, making them highly suitable for tasks like image recognition and classification [21]. CNNs use a sliding window approach to process local regions of the input data, allowing them to recognize patterns regardless of their position in the input. Pooling layers are often used to reduce spatial dimensions and control overfitting. In computer vision, CNNs have played a crucial role as they are well-suited for data processing with spatial hierarchies.

This review article covered very recent and prominent studies that implemented the specialized CNN architectures for medical imaging like U-Nets and others. The architectures that are well suited to the task of classification, detection and segmentation are discussed in much detail which not available in other literature reviews till mid of 2026. The reaming sections of the paper are described as follows: second section describes different modalities of images and the complexity

associated with them. Some reputed and popular datasets of medical images are also discussed in this section. Foundation and some application areas of CNN are discussed in 3rd and 4th section reactively. Section 5 is very important section that explain the task specific CNN architectures, followed some recent and important contributions of different studies in sixth section. Discussion of the various issues and challenges associated with implementation of different CNN architectures over the time has been described in section 7. The paper is concluded with some potential future directions at last.

TABLE I. KEY ASPECTS, DETAILS AND REFERENCES FOR CNNs IN MEDICAL IMAGING

Aspect	Description	References
Role of CNNs	CNNs are transformative in automated analysis and diagnosis in medical imaging, aiding faster and more accurate decision-making.	[1], [2], [3]
Applications	Applicable to X-rays, MRIs, CT scans, and histopathology slides; can detect anomalies, classify diseases, and predict treatment outcomes.	[3], [5], [6], [7], [8]
Mechanism	Utilize layers of artificial neurons and convolutional operations to identify complex patterns, structures, and anomalies within images.	[4], [5]
Advantages	Enhance precision and efficiency of image interpretation, enable early disease detection, accurate classification, and prediction of treatment outcomes.	[6], [7], [8]
Impact on Healthcare	Indispensable tools for healthcare providers, improve patient care, speed up diagnoses, and support personalized medicine.	[9], [10], [11], [14]
Significance in Medical Imaging	Revolutionized the traditional interpretation of medical images, reducing reliance solely on manual analysis by professionals.	[15], [17]
Technological Implications	Fusion of advanced technology with medical imaging is redefining healthcare, promoting data-driven approaches.	[12], [13]

II. Medical Image Modalities and Datasets

When it comes to providing high-quality treatment, medical imaging is essential across the board. Accurate diagnoses are crucial for making sound choices in public health, preventative medicine, curative treatment, and palliative care[13], [14]. Medical imaging division includes all diagnostic and therapeutic procedures that are performed in radiology department. It includes several methods of imaging the human body for diagnostic, therapeutic, and follow-up reasons, and it's crucial to efforts to boost community health. What's in it is:

- X-rays
- Computed tomography (CT)

- Ultrasound (US)
- MRI)
- Nuclear medicine: the use of radioactive chemicals in very small doses for medical purposes, including diagnosis, monitoring, and treatment.
- PET scans are used to identify and quantify patterns of radiopharmaceutical accumulation in the body after intravenous injection of a positron-emitting radiopharmaceutical is followed by a waiting period to allow for systemic dispersion.
- In the same way that PET provides metabolic and functional information, SPECT does the same for nuclear medicine [12].

The above mentioned variety of medical imaging may be acquired digitally, including retinal imaging, histology slides, and dermoscopy pictures. Multiple types of medical images are shown in Figure 1 . CT and MRI are two of the techniques mentioned above [6] that may evaluate many organs at once, whilst others, like retinal photography and dermoscopy, are specifically designed to examine a single organ. It is evident that the majority of imaging methods provide substantial volumes of data. While a histology slide may provide a limited number of pictures as data, an MRI has the potential to generate a much larger amount of data, sometimes reaching hundreds of gigabytes in terms of image size [22].

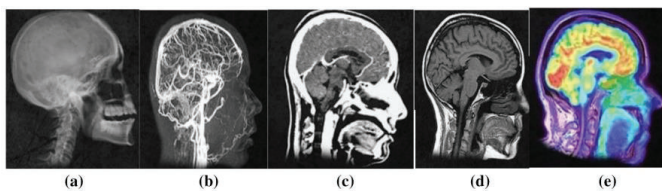


Fig. 1. (a) An axial X-ray of the brain with, (b) one-way MRA, (c) an MRI on the right side, (d) an axial CT brain scan (e) PET scan [23]

A. Necessity of Medical Imaging:

To record abnormalities in the human body, imaging methods are used. Accurate diagnosis, prognosis, and treatment planning for the abnormalities need comprehension of the collected pictures. Medical imaging should typically only be interpreted by qualified medical professionals. But the availability of human experts, their fatigue, and the requirement for approximative assessment techniques all restrict the effectiveness of human-done image interpretation in medicine. To comprehend images, CNNs are useful tools. They are so good at comprehending images that they have surpassed the performance of human specialists [17]. Diagnostic imaging services are essential for verifying, accurately diagnosing, & recording courses of many illnesses & for assessing responses to therapy, even though medical/clinical judgment is adequate before treatment for many ailments. When utilized correctly, imaging may be a great resource for physical therapists treating a wide variety of disorders. Imaging should only be performed when required; otherwise, resources would be wasted, and unrequired surgery might be performed too[13], [14].

B. Datasets:

Importance of an annotated medical dataset is a key concern while developing any CNN model as the success of the model is heavily dependent on the dataset. All the important tasks like classification, detection or segmentation of targeted lesion are possible only through the quality datasets. Medical datasets may be obtained in a variety of modalities viz. X-ray, CT scan, MRI, Ultrasound images, PET etc. A notable progress has been observed in recent past through the CNN models thanks the annotated quality datasets. A sound study of various types of datasets including brain , neck, chest a, abdomen and hematological images with a almost all modalities has been presented in [22]. Usually generic computer vision task involving thousands and millions of images as found in ImageNet (14 images), Microsoft COCO (2.5 labelled object instances), MNIST and CIFAR are used to trained very large model like AmoebaNet (55 million of parameters)[23]. However usually moderate to smaller datasets are available for medical images. MIMIC-CXR is relatively a larger datasets which contains 227,835 X-ray annotated images from 65,379 patients[24] whereas BraTS (Brain Tumor segmentation challenge) is a moderate size sample dataset which contains nearly 2000 cases . Being smaller in size BraTS is considered to be a complex dataset as it of diverse modalities of 3D MRI images of brain tumors along with class imbalance and uncertain boundary lines [25].

Analyzing medical image is much difficult task due to lots of factors including multimodality of the image, high class variance, low images sample availability and above it, lack of proper annotation. The complexity of the medical has been described in Figure 2 in a concise manner. The figure describes the very first challenge that is modality as a medical image may be a X-ray, MRI, CT scan , or PET. Further volumetric nature of data increase the complexity due multiple slices as shown in the figure. Another important issues are high variability and low contrast leading to identify a healthy tissue vs. diseased one among patients with their diverse physiology and anatomy. Artifact and noise make the process of tissue identification more difficult by reducing the image quality. As far as the classification task is concerned, rare positive class creates a much greater problem as compared to the balanced class distributions. In detection process the small lesion are often overlooked while localization process, whereas segmentation suffers from the poor outlines of the diseased tissue. Data scarcity and difference in expert opinion while marking the lesion are another significant overhead .Further, the heavy computation is involved in 3D or 4D like convolutions which demands high end systems, usually lead to memory shortage and slow training. Above all, generalization remains an open challenge like in all other deep learning models, as the image samples required for training are often limited and biased [5], [26].

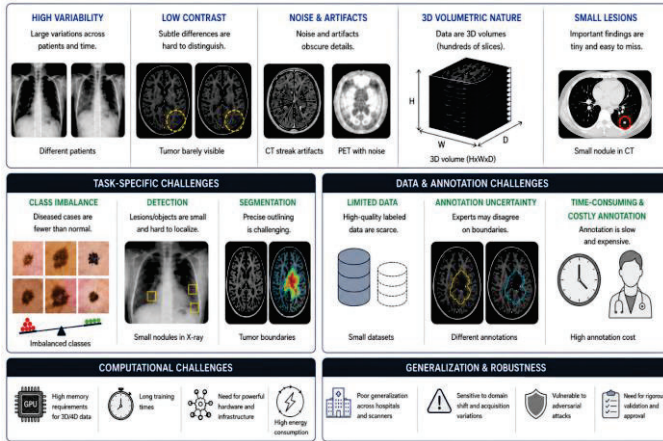


Fig. 2. Challenges across medical image analysis

III. Feed Forward and Convolutional Neural Networks

Before going into the details of CNN let a fundamental basis of all the neural network be presented in crisp manner. Feed forward neural networks (FFNNs) are general purpose neural architecture and can be used for a variety of machine learning tasks usually performed on tabular, text, and structured data in addition to image and video data. They are made up of stacked layers of neurons, with connections between all of the neurons in the preceding layer and those in the one above it. For FFNNs to understand intricate correlations in data, activation functions like ReLU and sigmoid are used to inject non-linearity into the model. Generic ML tasks like regression, classification, and function approximation are some applications of FFNNs. Different neural architecture like RNNs and LSTMs, are excellent in analyzing and processing structured data contains sequences. RNNs, which are fundamental building blocks for many deep learning applications, can also be applied to a wide variety of machine learning tasks like contemporary large language models.

A. Feed-Forward Neural Network (FFNN):

Multilayer Perceptron (MLP) is popular FFNN found to be most cited neural network in the literature of machine learning till beginning 21st century. In nutshell, FFNNs and MLPs are made up of layers of neurons, with connections established between all of the cells in the layer below. An input layer, many hidden layers, & an output layer make up the typical design. Mathematically, the operation of a single neuron in an FFNN can be represented as follows [21]:

$$a_i = \sigma(\sum_{j=1}^N w_{ij} x_j + b_i) \quad (1)$$

Where:

- a_i : output of neuron i.
- \sum : Summation operator prior to applying activation function

- w_{ij} : weights connecting neuron j in previous layer to neuron i
- σ : is an activation function like sigmoid or ReLU
- x_j : output of neuron j in previous layer.(output of the previous layer(ith node) neuron becomes input to the current layer neuron (jth node))
- b_i : bias term for neuron i.

B. Convolutional Neural Network:

CNNs (or ConvNet) are another class of FFNNs become much cited and popular in 21st century, are best suited for tasks involving grid-like data and excel at automatically learning hierarchical features. [21]. They gradually learn features from input data using convolutional layers with learnable filters and pooling layers. All CNNs have a convolutional layer, followed by a fully connected layer for classification or regression, in their hierarchical structure. Mathematically, operation of a convolutional layer in a CNN can be represented as follows [21]:

$$y_{i,j,k} = \sigma(\sum_{m=1}^M \sum_{n=1}^N \sum_{l=1}^L w_{m,n,l,k} x_{i+m-1,j+n-1,l} + b_k) \quad (2)$$

Where:

- $y_{i,j,k}$: output feature map at position (i, j) of channel k.
- $w_{m,n,l,k}$: weights of the convolutional filter for channel k, applied to the local region of the input.
- $x_{i+m-1,j+n-1,l}$: input value at position (i+m-1, j+n-1) in channel l.
- b_k : bias term for channel k.

FFNNs are more general and fully connected, making them suitable for a wide range of tasks, while ConvNets being specialized for grid-like data with spatial dependencies, such as images. Although the basic architecture of the CNN is obviously a FFNN type but the key difference is lying in their working and how they process data. CNNs are a common DL neural now become prevalent in an emerging field termed as Computer Vision. In summary we can say that ML is a field where ANNs really shine, computer vision is a field where the CNNs really shine while LLM is a field where an LSTM (a type of RNN) really shines [27].

1) Input Layers / Convolutional layers:

In this layer, users specify the actions they want their model to do. There are as many neurons in this layer as there are characteristics in the underlying data. For a two-dimensional input Image Im and output kernel Ke [21], we get the convolution process in the form of Equation (1).

$$S(x, y) = (Im * Ke)(x, y) = \sum_a \sum_b Im(a, b) Ke(x - a, x - b) \quad (3)$$

Where:

- "S(x, y)" represents the result of the convolution operation.
- "Im(a, b)" represents a two-dimensional discrete function (usually an image or a matrix) defined on a grid, with its values at discrete positions (a, b).
- "Kc(x-a, y-b)" represents a kernel or filter function, which is also defined on a grid and typically smaller in size than the input function "Im." This kernel is usually shifted across the input "Im" to perform local operations.

The summation symbols $\sum a$ and $\sum b$ indicate that you are summing over all possible values of "a" and "b" within the given bounds. This means you will perform the convolution operation for each possible position of the kernel relative to the input.

2) *Hidden Layer/ Activation layers:*

Data is sent from input layer to concealed layer. Intricacy of their model & quantity of data determines number of hidden layers. Each hidden layer often has more neurons than characteristics [2], [3]. In order to solve the issue of dying neurons, the ReLu was extended with new features, such as Leaky ReLu, PReLU, etc. [27].

$$ReLU = \max(0, x) \quad (4)$$

3) *Output Layer:*

Output of hidden layer is input into a logistic function like sigmoid or softmax to calculate probability score for each class. Equation (3) depicts output (out) of a neuron with input (i), weights (w), & bias (b) vectors. [27].

$$out = w^T i + b \quad (5)$$

To categorize the objects included inside an input image, a Convolutional Neural network (ConvNet/CNN) applies weights and biases to distinct regions of the image that it has learned. When compared to other classification algorithms, ConvNet needs much less setup time. While filters and characteristics must be hand-engineered in basic methods, ConvNets may learn them on their own with enough training.

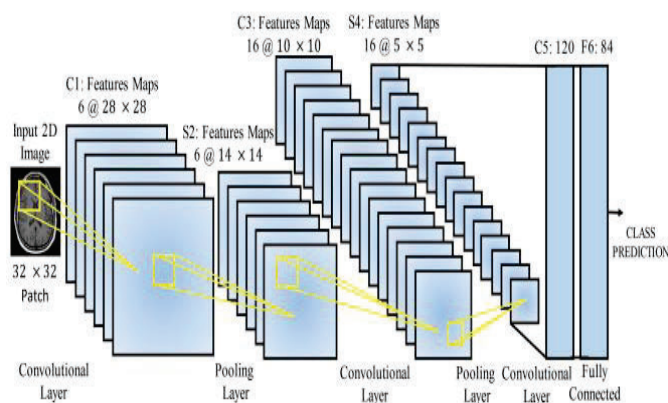


Fig. 3. A CNN structure for classification [20]

Medical image may be classified into N different categories using a CNN structure, as shown in Figure 3. The design relies on a 32x32 patch taken from the original 2D medical picture in order to function. Layers of convolution, max pooling, and fully linked nodes make up the network. The feature maps produced by the various convolutional layers are of varying sizes, and the pooling layers reduce these feature maps before passing them on. The classification prediction is delivered by fully linked layers at the output. The number of parameters required to adequately characterize a network is determined by the number of layers, the number of neurons in each layer, and the interconnections between neurons. The network's weights are optimized during training to provide the optimal response to the training challenge. As deep learning techniques and computing power have improved, researchers in the field of medical imaging have been motivated to apply them to the analysis of medical images. Recent studies have demonstrated that deep learning algorithms may be successfully used to a variety of medical tasks, including as image segmentation, CAD (Computer-Aided Diagnosis), disease categorization, and medical image retrieval. Predicting classes for relatively simple binary photos may yield an average accuracy score with this method. Still, it will certainly fail spectacularly when applied to more complex images with interdependent pixels.

IV. Applications of Convolution Neural Network

CNNs find applications in a wide range of fields and are particularly well-suited for tasks involving structured grid data, such as images and video. They have been shown to be effective in achieving accurate face detection in photographic pictures. The neural network receives an input image and outputs a set of numerical values that represent the face features and their respective positions within the image. The rapid and accurate identification of facial features, such as eyes, nose, & mouth, may be achieved with little distortion caused by perspective. CNNs and other imaging approaches have proven useful in the medical imaging area. They have been shown to improve the precision of X-ray and MRI cancer and abnormality identification. Images of particular human body parts, such the lungs, may be analyzed using CNN models to look for signs of cancer or other abnormalities [4], [5], [6].

Besides straight forward classification, detection and segmentation of medical images, CNNs are also a practical tool for document analysis. They have significant implications for recognition systems [7] and has relevance in the area of handwriting analysis. This makes the usage of CNN much effective in medical analysis since it will embed report generation also. In order to expeditiously examine text and compare it with an extensive database, a computer must process around one million instructions per minute. CNN networks have the capability to enhance document comprehension by using a combination of textual and graphical elements, hence facilitating the identification of relevant phrases [8].

CNNs have the capability to effectively capture and represent both visual pictures and the accompanying spatial information they include due to its superior capacity to extract features from pictures, CNNs are often considered as flexible non-linear function approximators. For example, the first layers of the network would capture basic features like edges, but subsequent layers will capture more complex attributes such as the contour of an object [9].

CNNs have been used for the purpose of biometric identification of individuals via the recognition of distinct facial traits. Convolutional neural network (CNN) models have the capability to undergo training using either photographs or videos containing individuals' facial images, enabling them to identify certain facial attributes such as interocular distance, nasal shape, and lip curvature [10]. CNN models have been successful in detecting and classifying many emotions shown by individuals in facial expressions captured in both photos and videos, including but not limited to emotions such as happiness and sorrow. Convolutional Neural Networks (CNNs) are more than capable of assessing the facial structure shown in frames of face images. [11].

V. CNN Models Used in Medical Image Analysis

Medical image analysis through CNN may contain disease diagnosis through classification, lesion detection and segmentation tasks. Although image registration, image enhancement, instance segmentation and image retrieval are other kind of tasks that may be performed by CNN but majorly classification, detection and segmentation are the key tasks for the diagnosis purpose. Various models have been proposed over the time for the above mentioned tasks. Although these tasks are related but they serve different purposes and therefore the CNN architecture for them also differs.

A. Architecture for classifying abnormality in medical images :

Classification determines the class of the abnormality or disease and as a result, the predicted label is assigned to the image under examination, the detection process identify the location of the abnormality and provide coordinates of the boundary box along with label. The segmentation is very important task in medical image analysis which outline the exact size and shape of the abnormality of any lesion like tumor or nodule. In this section the authors will be describing how the task specific CNN architecture emerged eventually. In the classification process if X represents the image and y represent the abnormality say 'malignant' or 'benign', then classification would be mathematical model f that maps X into y such that:

$$f: X \rightarrow y \text{ which actually means } \text{Image } X \rightarrow \text{benign}$$

The typical CNN architecture for the above task is pipeline:

Medical image \rightarrow Convolution layer \rightarrow Pooling layer \rightarrow FC layer \rightarrow output as image label

The popular architectures for the classification are Visual Geometry Group (VGG) variants like VGG 16 and VGG 19 , ResNet, AlexNet, DenseNet, EfficientNet ,GoogleNet and many more [13]. The steps of the classification mentioned in the above pipelined have already benn explained in Fig. 3 presented in 3rd section.

Firstly images of a diverse modality are preprocessed via resizing, data augmentation and denoising prior to pass to convolutional layer. As per the popular text book "Deep Learning", by Ian Goodfellow (2016), a convolutional layer involves convolution, detector and pooling steps as a combined big step to extract features to form a feature map [28]. At last fully connected (FC) layer perform classification and ultimately disease diagnosis through label prediction. It is important to note that over the period the depth of the CNNs went on increasing. For example a DenseNet network is much deeper than ResNet, which is itself deeper than an AlexNet network . The AlexNets have sequential architecture, the ResNets have residual (skip) connections while DenseNets have dense connection making them more potent and viable for classification. Today a lot of intense deep and extensive parameterized CNNs can be observed which are used for medical image classification[29].

B. Architectures for lesion detection:

Lesion detection is another popular and important task which has greater significance in real time scenario, however its use in offline mode is equally popular in medical imaging. Lesion, diseased tissue, noodle or tumor identification with localization and positional details, are obtained through detection mechanism which makes it better than merely classifying the abnormality in the image. The initial thought of lesion detection came with proposal of R-CNN (Region based CNN) which answer the question where the abnormality is lying by localization with boundary box. In classification sliding window process is exhaustive one, whereas in detection a selective search is made by the window so as to find the region of interest among the candidate regions. It is obvious that a selected candidate region will be an imprecise region so the boundary box regression is performed to refine the lesion boundary. Finally Non-Maximum Suppression (NMS) is applied to select the region with most confidence. The cost of detection was often high in R-CNN, that has been optimized and made efficient later on via Fast R-CNN, Faster R-CNN and Mask R-CNN [30]. For real time detection, single stage detection model like Yolo and its variants have been proposed over the time[31], [32].

C. Architectures for medical image segmentation:

Medical image segmentation is most desired analysis on that should be performed on an image of arbitrary modality. Each pixels must be labeled in order to segment the image unlike classification whereas a single label is sufficient for the entire image. It is also more precise than detection as it provides exact delineation of the effected part or lesion. Diagnosis become eventually effective and accurate when the

actual delineation tumor, diseased tissue, or noddled under examination. A variety of architectures have been over period for the segmentation including V-Net , U-Net and its variants like U-Net++ , attention U-Net, TransU-Net, SwinUNETR and Mamab-UNet [33], [34] . A detailed discussion of various CNNs specifically utilized for segmentation is beyond the scope of this paper, however the two representative popular architecture for same have been describe below: U-Net and TransUnet.

D. U-Net architecture:

U-Net CNN proposed by Ronneberger et al. in 2015 with purpose of biomedical image segmentation and they claimed it very successful due its unique architecture especially skip connection as shown in the Fig.4. The brain MR image has been shown as input so as to exemplify the role of U-Net in medical image analysis. The idea of contractive and expansive path corresponds to encoding and decoding respectively. The expansive path, applies the modules to downscale the image by implementing several blocks of convolution for extracting the semantic and contextual features. On other hand expansive path implements the up-scaling of the feature map by applying the convolutional blocks repeatedly to produce sufficient resolution in the feature map. Skip connections keep the high level of understanding and local details of the image unchanged by copying the output at each stage [34], [35]. The discussion of the U-Net architecture was mandatory as it was breakthrough in journey of medical image analysis via CNNs.

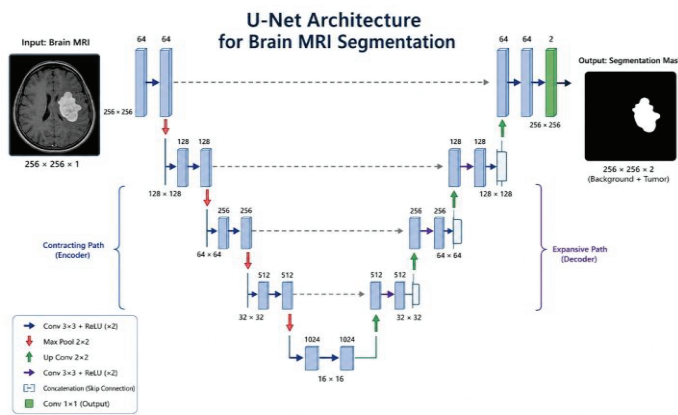


Fig. 4. U-Net architecture for medical image segmentation for a Brain MRI image

VI. Literature Review Describing the Role of CNNs in Medical Imaging

The literature review presented here covers most of the image related tasks however classification and segmentation will remain the core. Medical image classification is the classic application of deep learning whereas image segmentation is much desired as it not sufficient to only identify the abnormality in the image. In fact it is much needed that image should be localized, well separated from background and the volume of the abnormality should be properly known for its actual risk assessment and

accurate diagnosis especially brain tumor like diseases. U-net found to be most successful in the segmentation task as it pass through, down sampling, first, bottleneck and up sampling at last. A thorough review has been done by a very recent studies that utilized advanced U-Net variants for glioma, a brain tumor, in MRI images[36].

Since U-Net assumed to be the very first successfully customized CNN model for medical image analysis, an enhancement in U-net is realized a study in 2026, where three pipelines are proposed. In the proposed pipelines the MRI images are first preprocessed via CLAHE, MBOBHE and MPHE separately, then the preprocessed image is fed to the network through U-net encoder, Vision Transformer and then U-Net Decoder so as to perform segmentation mask ultimately with accuracy 99%, sensitivity 99%, specificity 100% and with AUC 1 for the dataset BraTS2020[37]. The standard U-Net architecture has been augmented with inception decoder and transfer learning in a novel study to segment the tumor in breast ultrasound images[38]. Overall U-Net models are the successful in identifying tumors but with reasonably good accuracies and they have been modified with transformer and attention mechanism over the time [39]. In review study of 2022, a thorough overview of CNN's usefulness in the field of medical image analysis was discussed. The fundamental goal is to encourage the use of CNNs in medical image comprehension research and diagnostics. This article served as a primer that surveyed 3D CNNs for medical image analysis especially segmentation of the images. Many excellent CNN frameworks have been discussed [18], [19]. Research has presented the four main tasks involved in medical picture understanding: image classification, segmentation, localization, and detection [20].

CNNs may be in many classification tasks involving sound and images. For two-dimensional information, CNN performs well. It has numerous convolutional layers, fully linked layers, pooling layers, and normalization layers among its many hidden layers.

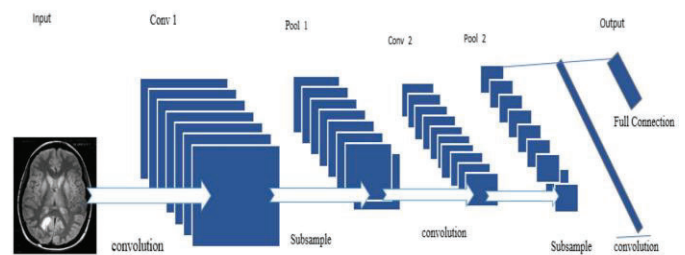


Fig 4. The architecture of convolutional neural network using medical image [6]

In the field of image processing, particularly for the purpose of photo categorization, several CNN approaches have been created employing deep learning techniques. The difficulty of medical image processing is dramatically heightened in the absence of continual breakthroughs and expansion of convolutional neural networks. Simplified CNN architecture with example representative features and convolution and pooling processes is shown in Figure 4. These operations are

employed to extract representative features and establish relationships between pixels in the input pictures. The variability in the complexity of features in trained kernels may be seen. At the outset, superficial characteristics such as edges are removed. Subsequently, intricate and advanced features are extracted.

R. Mohakud et al. (2022) considered an automated hyper-parameter optimized convolutional neural network to identify the kind of skin cancer. The approach has optimized the hyper parameters of CNN using Grey Wolf Optimization and a suitable encoding strategy. Using the skin lesion multi class data set from the International Skin Imaging Collaboration, the model's efficacy is validated by comparing the results to those obtained using a hyper-parameter optimized CNN based on Particle Swarm Optimization and a Genetic algorithm. Compared to PSO and GA based models, simulation studies reveal that the recommended model can reach a testing accuracy of up to 98.33 percent. The proposed model achieves a testing loss of around 0.17%, which is comparable to PSO and GA based models and is 39.2% and 15% lower, respectively. The experimental findings prove without a shadow of a doubt that the suggested model can hold its own against the other published models [1].

M. J. Lim et al. (2018) focused on the use of a deep CNN to analyze medical images. Many individuals all throughout the globe have died from breast cancer since it was so easily spread. If caught early enough, it may be restored completely. It was crucial to correctly identify whether or not a tumor was breast cancer to facilitate early identification. Histopathology pictures of breast cancer have recently been used to demonstrate the superior accuracy and efficiency of the deep learning approach technique over more traditional approaches. This research presents a visual analysis of the histological appearance of breast cancer that is difficult to differentiate. VGG16 and InceptionV3 were employed as CNN algorithms, with transfer learning applied to make the most of them [2].

A. Voulodimos et al. (2018) introduced the recent meteoric rise of deep learning may be attributed in large part to the advancements it has made possible in the field of computer vision. These three main types of DL for computer vision—CNNs, the "Boltzmann family" (DBNs and DBMs), and SdAs (Stacked Denoising Autoencoder)—have been used to achieve significant performance rates in a wide range of visual understanding tasks, including object detection, face recognition. However, there are perks and drawbacks to every group. In contrast to other models, CNNs may learn features automatically depending on the dataset provided. Another remarkable feature of CNNs for certain computer vision tasks is their invariance to transformations [3].

CNN is a common model structure in machine learning, and a 2016 evaluation by K. Masaoka et al. concluded that it yields the best results when used to image enhancement, image fusion, IP (Image Processing), image recognition, and so on. When applied to the initial CNN model, the data from the rejected images helps to boost the accuracy with which machines can learn and classify new data. In the case of image

recognition in particular, the machine can correctly identify the picture feature information following several operations, including the computation, pooling, and abstraction of CNN. There will be a lot of progress made in the field of machine vision in the years to come [4]. M. Puttagunta and S. Ravi (2021) spoke about recent advances in deep learning algorithms for medical imaging systems. The researchers looked at the AI's ability to detect anomalies in magnetic resonance imaging (MRI), computed tomography (CT) scans, X-rays and ultrasounds. The scientists examined different learning models to demonstrate how deep learning might help in early prediction of diseases, reduce the human effort and increase the accuracy. The essay also discussed the issue of overfitting, training data quality and intelligent system integration into real time healthcare setting[4].

J. Ker, L. Wang, J. Rao and T. Lim (2017) provided a comprehensive overview of deep learning techniques in medical image processing. The research showed that Convolutional Neural Networks are beneficial to recognize diseases, classify and segment them and give diagnostic support. The authors pointed out the advantages of automatic feature extraction in comparison to more conventional methods e.g., better accuracy. It also highlighted issues such as computational complexity, limited datasets and clinical validation [5]. S. S. Kshatri et al. (2023) gave a comprehensive summary of CNN medical image analysis, including its strengths and weaknesses. Use of CNNs in medical image analysis, particularly in pre- and post-processing, segmentation, data preparation, and MRIs of the brain, was of major interest. Further, they investigated how large-scale retrieval might improve the effectiveness of MRI image processing. In sum, this research examined both recent successes and formidable obstacles in this field. Before categorising the papers using a taxonomy based on human architecture, they were sorted first by pattern recognition[6].

The results of studies by X. Song et al. (2022) on the use of DL techniques to the analysis of medical pictures are presented. It was found that CNN-based DL algorithms are being applied increasingly often in all fields of medical image processing[7]. S. Kilicarslan et al. (2020) focused on the convergence to the global optimum was possible via SGD. WKCNN contains deep layers and too many parameters, making it vulnerable to local optimization if the parameters are not provided correctly. As a result, the Adam approach (Kingma and Ba, 2014) is used all through this analysis. Adam's resistance to parameter selection makes him a handy tool for configuring neural network parameters[8]. N. Noreen et al. (2020) reviewed the deep learning models for tumor diagnosis in the brain was described in this research. In this research, research compare two possible futures and provide their findings. The characteristics used in the study originated in a pre-trained deep learning network, namely DensNet201. Then, a softmax classifier was used to classify the kind of brain tumour based on the combined criteria. The second step included taking features from several pre-trained Inception module[9].

P. Sharma et al. (2022) proposed brain tumor classification using the InceptionV3 model combined with a softmax classifier. Both scenarios were put to the test using data on three distinct kinds of brain tumours that was made available to the general public. The proposed method tested the best in identifying brain tumors, with a success rate of 99.51 percent using sample cases. Researchers want to improve current methods for detecting brain tumors by using fine-tune techniques to pre-trained models trained with a greater number of layers and by building new models from scratch using data augmentation approaches. The ensemble method will also be studied by gathering deep learning models' fine-tuned and scratch-based characteristics[10].

L. Abdelrahman et al. (2021) reviewed the gaps in smaller, less annotated datasets; researchers have turned to transfer learning. More research was necessary to see whether this was applicable to the problem of detecting breast asymmetry. More annotated datasets are required for this method; hence future research should focus on creating large corpora. As a result of their investigation, research has compiled a comprehensive list of publicly available breast picture datasets[11]. Y. D. Zhang et al.(2021) focused on the everyone in society has access to high-quality medical care, thanks to the healthcare system. It also conducts studies to discover and counteract novel viruses and other illnesses. A healthy lifestyle is essential for everyone since one's level of health is a major influence in determining one's potential. The study of medicine helps satisfy this universal need for physical well-being. In today's fast-paced, technologically-advanced society, it was vital and unavoidable to incorporate new technologies into healthcare. Health care processes including diagnosis, treatment, medication prescription, etc., may be streamlined and expedited with the use of modern technology. The necessity for a skilled workforce is also diminished. As a result, it is essential to move medical research forward to utilize diverse technologies in healthcare. In this study, research takes a look at how several

ANN-based methods might help with breast cancer diagnosis [12].

H. Yu et al. (2021) introduced the CNN in the study of medical images. As a kind of deep learning model, CNN were widely used. Recent advances in computer-aided diagnosis may be attributed in large part to the widespread usage of CNNs in many applications of medical image processing. In this study, research present a comprehensive review of the use of CNN for this purpose. First, research take a look back at the history of CNNs and how they've been put to use in the field of medical imaging. They next provide an overview of how CNNs were being used in common medical diagnostic domains, including the brain, breast, and abdomen, to perform tasks including classification, segmentation, detection, registration, content-based image retrieval, picture production, and enhancement. Finally, they outline the remaining difficulties in using CNNs for medical image processing and suggest potential avenues for further study [13]. The suggested method for medical image fusion based on CNN and non-sub sampled contourlet transform examined the source multi-modality photos that were separated into low and high frequency sub bands. Ten of the best MIF (Medical Image Fusion) algorithms already available are chosen to serve as benchmarks for the proposed approach. The suggested approach outperforms the other comparable algorithms in fusing multimodal medical pictures, as shown by subjective assessments from five physicians and objective evaluations from seven image quality criteria [15]. CNNs are regarded efficient for segmenting medical pictures by Z. Han et al. (2022). Their model has 20% less parameters than the standard U-Net. The experimental results on many datasets show that it has enhanced segmentation performance regardless of data size[15].

A summary of the selected studies has been presented in the Table 2 for a consolidated view of the CNNs journey till date.

TABLE II. LITERATURE REVIEW

Author(s) & Year	Focus / Objective	Methodology / Model	Dataset	Key Findings / Results
R. Mohakud et al. (2022)	Skin cancer classification	Hyper-parameter optimized CNN using Grey Wolf Optimization (GWO)	ISIC skin lesion multi-class dataset	Testing accuracy up to 98.33%, testing loss ~0.17%; outperforms PSO and GA-based models [1]
M. J. Lim et al. (2018)	Breast cancer detection	Deep CNN (VGG16 & InceptionV3) with transfer learning	Histopathology images	Superior accuracy & efficiency compared to traditional methods [2]
A. Voulodimos et al. (2018)	DL in computer vision	CNNs, DBNs, DBMs, SdAs	Various CV datasets	CNNs learn features automatically and are transformation-invariant [3]
M. Puttagunta et al. (2021)	Medical image analysis	Review of supervised & unsupervised DL (CNN, RNN, auto-encoders, RBMs)	Medical images	Overview of DL algorithms for clinical decision support [4]
J. Ker et al. (2017)	Medical image analysis	Review of DL methods & frameworks (Caffe, TensorFlow, PyTorch)	Medical images	Comprehensive overview of DL applications in clinical imaging [5]
S. S. Kshatri et al. (2023)	MRI & medical image processing	CNN for pre/post-processing, segmentation, large-scale retrieval	MRI datasets	Identified strengths, weaknesses, and challenges in CNN-based medical imaging [6]

X. Song et al. (2022)	DL in medical image analysis	CNN-based DL algorithms	Various medical image datasets	CNN-based methods increasingly applied in medical imaging [7]
S. Kilicarslan et al. (2020)	Neural network parameter optimization	WKCNN with Adam optimizer	Medical images	Adam improves parameter tuning, avoids local optima [8]
N. Noreen et al. (2020)	Brain tumor diagnosis	DenseNet201, pre-trained Inception modules, softmax classifier	Brain tumor images	Accurate tumor classification using DL features [9]
P. Sharma et al. (2022)	Brain tumor classification	Concatenated V3 model + softmax classifier	Public brain tumor datasets	Success rate 99.51%; ensemble & fine-tuning recommended [10]
L. Abdelrahman et al. (2021)	Breast image analysis	Transfer learning	Breast image datasets	Emphasis on need for large annotated datasets[11]
Y. D. Zhang et al. (2021)	Healthcare automation & ANN	ANN-based methods for breast cancer	Medical datasets	Modern technology streamlines diagnosis & treatment [12]
H. Yu et al. (2021)	CNN in medical imaging	Comprehensive review of CNN applications	Brain, breast, abdomen images	Tasks: classification, segmentation, detection, registration, enhancement; challenges & future directions outlined [13]
Unspecified (Medical Image Fusion)	Multi-modal medical image fusion	CNN + Non-Subsampled Contourlet Transform (NSCT)	Multi-modal medical images	Outperforms top 10 existing MIF algorithms; validated by physician assessment and quality metrics [14]
Z. Han et al. (2022)	Medical image segmentation	CNN with 20% fewer parameters than UNet	Various segmentation datasets	Improved segmentation performance regardless of dataset size [15]

VII. Discussion

Analyzing medical images with CNNs presents a host of unique challenges and issues. One of the foremost concerns is small availability of healthcare information, especially labeled datasets, which can be challenging and costly to acquire. Additionally, medical datasets often suffer from class imbalance, where certain diseases or conditions are rare, leading to biased models. Ensuring the quality of annotations in medical image datasets is crucial, as inaccuracies can detrimentally affect model training and evaluation. Interpretability and explainability of CNN decisions are paramount in medical applications, as understanding why a model made a particular diagnosis is critical for clinical acceptance. Furthermore, the generalization of CNNs across different medical datasets can be problematic due to variations in equipment, patient demographics, and acquisition protocols. Robustness to noise in medical images is also essential for reliable diagnoses. Ethical considerations with record of patient privacy compliance, add complexity to data acquisition and model deployment. Additionally, clinical validation and regulatory approvals are essential for ensuring safety & efficacy of CNN-based diagnostic tools in real-world healthcare settings. Adversarial attacks pose a security risk, and hardware and deployment challenges must be addressed to integrate CNN models into clinical workflows. Combining deep learning expertise with medical domain knowledge is crucial, and models need to be continually updated to adapt to evolving medical knowledge and technology. Despite these challenges, CNNs have made significant strides in healthcare image analysis, offering valuable support in detection of disease before segmentation, and prognosis prediction when appropriately navigated by multidisciplinary teams of experts.

The necessity for research in the domain of using CNNs to analyze medical pictures is crucial and motivated by numerous compelling factors. A broad variety of illnesses and conditions may be detected, diagnosed, and monitored with the use of medical imaging, making it an essential part of healthcare. CNNs shown remarkable potential in automating and enhancing the accuracy of image interpretation, potentially reducing diagnostic errors and improving patient outcomes. One crucial area of research is development of highly specialized CNN architectures optimized for different medical imaging modalities, such as radiography, magnetic resonance imaging, & pathology slides. These tailored models can improve the precision of disease detection, aid in early diagnosis, and facilitate more effective treatment planning. Moreover, addressing data-related challenges remains a pressing concern. The scarcity of labeled medical image datasets, the presence of class imbalances, and the need for reliable annotations necessitate research into data augmentation techniques. This research is crucial to ensure that CNNs can generalize effectively across diverse patient populations and clinical settings.

VIII. Conclusion and Scope of research work

This research provides an in-depth review of how CNN are being used as main tool for analysis of the medical images especially multimodal images. The research carried out begins with a focus on studying the evolution of CNNs. Further it explains how CNNs utilized in medical imaging and how they have grown over the time. This study then offers an overview of how CNNs are helpful in medical diagnosis, which helps in classification, segmentation, and detection primarily. However registration, content-based image retrieval, picture generation, and enhancement of the images (of brain, breast, and abdomen

especially) are also discussed, though not in that much details as the complete coverage is beyond the scope of this review. Almost all challenges usually arise while utilizing CNNs in medical image processing are discussed. The CNN based architectures have offered new avenues in diagnosis and treatment of the patients. Researches in the medical imaging field like radiology are now need to be tuned with CNN-based solutions.

Scope of Research:

In spite of CNN's great success in medical imaging especially in classification, detection and segmentation, some challenges are still open and need to be explored like generalization, reliability, AI-explainability and efficiency in real time systems. The most of the portions of a medical image are correlated spatially and having long term dependencies, so it is difficult even for transformer based CNNs to analyses them accurately. A much needed research area known as multi-modal fusion of the images of diverse modality like MRI, PET, and CT scan etc. is expected for the accurate detection via CNNs.

REFERENCES

- [1] R. Mohakud and R. Dash, "Designing a grey wolf optimization based hyper-parameter optimized convolutional neural network classifier for skin cancer detection," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, pp. 6280–6291, Sep. 2022, doi: 10.1016/j.jksuci.2021.05.012.
- [2] M. J. Lim, D. E. Kim, D. K. Chung, H. Lim, and Y. M. Kwon, "Deep Convolution Neural Networks for Medical Image Analysis," *Int. J. Eng. Technol.*, vol. 7, no. 3.33, pp. 115–119, Aug. 2018, doi: 10.14419/ijet.v7i3.33.18588.
- [3] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep Learning for Computer Vision: A Brief Review," *Comput. Intell. Neurosci.*, vol. 2018, no. 1, p. 7068349, 2018, doi: 10.1155/2018/7068349.
- [4] M. Puttagunta and S. Ravi, "Medical image analysis based on deep learning approach," *Multimed. Tools Appl.*, vol. 80, no. 16, pp. 24365–24398, Jul. 2021, doi: 10.1007/s11042-021-10707-4.
- [5] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep Learning Applications in Medical Image Analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2018, doi: 10.1109/ACCESS.2017.2788044.
- [6] S. Kshatri and D. Singh, "Convolutional Neural Network in Medical Image Analysis: A Review," *Arch. Comput. Methods Eng.*, vol. 30, Mar. 2023, doi: 10.1007/s11831-023-09898-w.
- [7] X. Song, Y. Cong, Y. Song, Y. Chen, and P. Liang, "A bearing fault diagnosis model based on CNN with wide convolution kernels," *J. Ambient Intell. Humaniz. Comput.*, vol. 13, no. 8, pp. 4041–4056, Aug. 2022, doi: 10.1007/s12652-021-03177-x.
- [8] S. Kiliarslan, K. Adem, and M. Celik, "Diagnosis and classification of cancer using hybrid model based on ReliefF and convolutional neural network," *Med. Hypotheses*, vol. 137, p. 109577, Apr. 2020, doi: 10.1016/j.mehy.2020.109577.
- [9] N. Noreen, S. Palaniappan, A. Qayyum, I. Ahmad, M. Imran, and M. Shoaib, "A Deep Learning Model Based on Concatenation Approach for the Diagnosis of Brain Tumor," *IEEE Access*, vol. 8, pp. 55135–55144, 2020, doi: 10.1109/ACCESS.2020.2978629.
- [10] P. Sharma and A. P. Shukla, "Brain Tumor Classification Using Convolution Neural Network," in *Proceedings of International Conference on Recent Trends in Computing*, R. P. Mahapatra, S. K. Peddoju, S. Roy, P. Parwekar, and L. Goel, Eds., Singapore: Springer Nature, 2022, pp. 579–588. doi: 10.1007/978-981-16-7118-0_50.
- [11] L. Abdelrahman, M. Al Ghamdi, F. Collado-Mesa, and M. Abdel-Mottaleb, "Convolutional neural networks for breast cancer detection in mammography: A survey," *Comput. Biol. Med.*, vol. 131, p. 104248, Apr. 2021, doi: 10.1016/j.compbiomed.2021.104248.
- [12] Y.-D. Zhang, S. C. Satapathy, D. S. Guttery, J. M. Górriz, and S.-H. Wang, "Improved Breast Cancer Classification Through Combining Graph Convolutional Network and Convolutional Neural Network," *Inf. Process. Manag.*, vol. 58, no. 2, p. 102439, Mar. 2021, doi: 10.1016/j.ipm.2020.102439.
- [13] H. Yu, L. T. Yang, Q. Zhang, D. Armstrong, and M. J. Deen, "Convolutional neural networks for medical image analysis: State-of-the-art, comparisons, improvement and perspectives," *Neurocomputing*, vol. 444, pp. 92–110, Jul. 2021, doi: 10.1016/j.neucom.2020.04.157.
- [14] Z. Wang, X. Li, H. Duan, Y. Su, X. Zhang, and X. Guan, "Medical image fusion based on convolutional neural networks and non-subsampled contourlet transform," *Expert Syst. Appl.*, vol. 171, p. 114574, Jun. 2021, doi: 10.1016/j.eswa.2021.114574.
- [15] Z. Han, M. Jian, and G.-G. Wang, "ConvUNeXt: An efficient convolution neural network for medical image segmentation," *Knowl.-Based Syst.*, vol. 253, p. 109512, Oct. 2022, doi: 10.1016/j.knsys.2022.109512.
- [16] D. Müller, I. Soto-Rey, and F. Kramer, "An Analysis on Ensemble Learning Optimized Medical Image Classification With Deep Convolutional Neural Networks," *IEEE Access*, vol. 10, pp. 66467–66480, 2022, doi: 10.1109/ACCESS.2022.3182399.
- [17] D. Müller, I. Soto-Rey, and F. Kramer, "An Analysis on Ensemble Learning Optimized Medical Image Classification With Deep Convolutional Neural Networks," *IEEE Access*, vol. 10, pp. 66467–66480, 2022, doi: 10.1109/ACCESS.2022.3182399.
- [18] S. Niyas, S. J. Pawan, M. Anand Kumar, and J. Rajan, "Medical image segmentation with 3D convolutional neural networks: A survey," *Neurocomputing*, vol. 493, pp. 397–413, Jul. 2022, doi: 10.1016/j.neucom.2022.04.065.
- [19] A. Sabeeh Yousif, Z. Omar, and U. Ullah Sheikh, "An improved approach for medical image fusion using sparse representation and Siamese convolutional neural network," *Biomed. Signal Process. Control*, vol. 72, p. 103357, Feb. 2022, doi: 10.1016/j.bspc.2021.103357.
- [20] M.-L. Huang and Y.-Z. Wu, "Semantic segmentation of pancreatic medical images by using convolutional neural network," *Biomed. Signal Process. Control*, vol. 73, p. 103458, Mar. 2022, doi: 10.1016/j.bspc.2021.103458.
- [21] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, "Medical Image Analysis using Convolutional Neural Networks: A Review," *J. Med. Syst.*, vol. 42, no. 11, p. 226, Oct. 2018, doi: 10.1007/s10916-018-1088-1.
- [22] J. Li et al., "A Systematic Collection of Medical Image Datasets for Deep Learning," *ACM Comput. Surv.*, vol. 56, no. 5, pp. 1–51, May 2024, doi: 10.1145/3615862.
- [23] E. Real, A. Aggarwal, Y. Huang, and Q. V. Le, "Regularized evolution for image classifier architecture search," in *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence*, in *AAAI'19/IAAI'19/EAAI'19*, vol. 33, Honolulu, Hawaii, USA: AAAI Press, Jan. 2019, pp. 4780–4789. doi: 10.1609/aaai.v33i01.33014780.
- [24] A. E. W. Johnson et al., "MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports," *Sci. Data*, vol. 6, no. 1, p. 317, Dec. 2019, doi: 10.1038/s41597-019-0322-0.
- [25] H. B. Li et al., "The Brain Tumor Segmentation (BraTS) Challenge 2023: Brain MR Image Synthesis for Tumor Segmentation (BraSyn)," *ArXiv*, p. arXiv:2305.09011v6, Nov. 2024.
- [26] A. W. Salehi et al., "A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope," *Sustainability*, vol. 15, no. 7, p. 5930, Mar. 2023, doi: 10.3390/su15075930.
- [27] P. Bir and V. E. Balas, "A Review on Medical Image Analysis with Convolutional Neural Networks," in *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)*, Oct. 2020, pp. 870–876. doi: 10.1109/GUCON48875.2020.9231203.
- [28] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*. The MIT Press, 2016. Accessed: Jun. 22, 2026. [Online]. Available: <https://www.deeplearningbook.org/>

- [29] I. D. Mienye, T. G. Swart, G. Obaido, M. Jordan, and P. Ilono, "Deep Convolutional Neural Networks in Medical Image Analysis: A Review," *Information*, vol. 16, no. 3, p. 195, Mar. 2025, doi: 10.3390/info16030195.
- [30] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.
- [31] M. G. Ragab et al., "A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023)," *IEEE Access*, vol. 12, pp. 57815–57836, 2024, doi: 10.1109/ACCESS.2024.3386826.
- [32] T. Feng, Y. Shi, X. Wang, H. Zhao, and W. Chao, "YOLO-MFDS: Medical small object detection algorithm based on multi-feature fusion," *Biomed. Signal Process. Control*, vol. 115, p. 109410, Apr. 2026, doi: 10.1016/j.bspc.2025.109410.
- [33] E. Benedykciuk, M. Denkowski, and G. M. Wójcik, "Differentiable Neural Architecture Search for medical image segmentation: A systematic review and field audit," *Comput. Med. Imaging Graph.*, vol. 128, p. 102713, Feb. 2026, doi: 10.1016/j.compmedimag.2026.102713.
- [34] R. Azad et al., "Medical Image Segmentation Review: The Success of U-Net," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 10076–10095, Dec. 2024, doi: 10.1109/TPAMI.2024.3435571.
- [35] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," May 18, 2015, arXiv: arXiv:1505.04597. doi: 10.48550/arXiv.1505.04597.
- [36] C. Jin and H. Ibrahim, "Advancements in brain tumor segmentation: a literature survey of U-Net variants," *Neural Comput. Appl.*, vol. 38, no. 5, p. 110, Mar. 2026, doi: 10.1007/s00521-025-11798-y.
- [37] S. Xinxin et al., "Global-local feature fusion in MRI brain tumor segmentation via enhanced U-Net-ViT architecture and adaptive contrast preprocessing," *Vis. Comput.*, vol. 42, no. 1, p. 125, Jan. 2026, doi: 10.1007/s00371-025-04241-9.
- [38] Y. Choi, M. N. Kim, and S. Na, "Inception U-Net for Enhanced Breast Ultrasound Image Segmentation Using Transfer Learning," *Bioengineering*, vol. 13, no. 2, p. 181, Feb. 2026, doi: 10.3390/bioengineering13020181.
- [39] M. Benchari, M. W. Totaro, M. Bayoumi, and S. Mokhtari, "The U-Net Architecture for MRI Brain Tumor Segmentation, Detection, and Classification: A Survey," *Biomed. Mater. Devices*, Feb. 2026, doi: 10.1007/s44174-026-00660-x.