



Review

# Medical Image Classifications Using Convolutional Neural Networks: A Survey of Current Methods and Statistical Modeling of the Literature

Foziya Ahmed Mohammed<sup>1,2,3</sup>, Kula Kekeba Tune<sup>1,2,\*</sup> , Beakal Gizachew Assefa<sup>4</sup> , Marti Jett<sup>5</sup> and Seid Muhie<sup>6,7,\*</sup>

<sup>1</sup> Department of Software Engineering, College of Electrical and Mechanical Engineering, Addis Ababa Science and Technology University, Addis Ababa 16417, Ethiopia; foziya.ahmed@aastu.edu.et

<sup>2</sup> Center of Excellence for HPC and Big Data Analytics, Addis Ababa Science and Technology University, Addis Ababa 16417, Ethiopia

<sup>3</sup> Department of Information Technology, College of Computing and Informatics, Wolkite University, Wolkite P.O. Box 07, Ethiopia

<sup>4</sup> School of Information Technology and Engineering, Addis Ababa Institute of Technology, Addis Ababa University, Addis Ababa P.O. Box 1000, Ethiopia; beakal.gizachew@aait.edu.et

<sup>5</sup> Head Quarter, Walter Reed Army Institute of Research, Silver Spring, MD 20910, USA; marti.jett-tilton.civ@health.mil

<sup>6</sup> Medical Readiness Systems Biology, Walter Reed Army Institute of Research, Silver Spring, MD 20910, USA

<sup>7</sup> The Geneva Foundation, Silver Spring, MD 20910, USA

\* Correspondence: kula.kekeba@aastu.edu.et (K.K.T.); smuhie@genevausa.org (S.M.); Tel.: +251-92-941-6857 (K.K.T.); +1-314-6171 (S.M.)

**Abstract:** In this review, we compiled convolutional neural network (CNN) methods which have the potential to automate the manual, costly and error-prone processing of medical images. We attempted to provide a thorough survey of improved architectures, popular frameworks, activation functions, ensemble techniques, hyperparameter optimizations, performance metrics, relevant datasets and data preprocessing strategies that can be used to design robust CNN models. We also used machine learning algorithms for the statistical modeling of the current literature to uncover latent topics, method gaps, prevalent themes and potential future advancements. The statistical modeling results indicate a temporal shift in favor of improved CNN designs, such as a shift from the use of a CNN architecture to a CNN-transformer hybrid. The insights from statistical modeling point that the surge of CNN practitioners into the medical imaging field, partly driven by the COVID-19 challenge, catalyzed the use of CNN methods for detecting and diagnosing pathological conditions. This phenomenon likely contributed to the sharp increase in the number of publications on the use of CNNs for medical imaging, both during and after the pandemic. Overall, the existing literature has certain gaps in scope with respect to the design and optimization of CNN architectures and methods specifically for medical imaging. Additionally, there is a lack of post hoc explainability of CNN models and slow progress in adopting CNNs for low-resource medical imaging. This review ends with a list of open research questions that have been identified through statistical modeling and recommendations that can potentially help set up more robust, improved and reproducible CNN experiments for medical imaging.

**Keywords:** medical imaging; convolutional neural network models; classification; hyperparameter tuning; frameworks; preprocessing; performance metrics; ensemble; activation function



**Citation:** Mohammed, F.A.; Tune, K.K.; Assefa, B.G.; Jett, M.; Muhie, S. Medical Image Classifications Using Convolutional Neural Networks: A Survey of Current Methods and Statistical Modeling of the Literature. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 699–735. <https://doi.org/10.3390/make6010033>

Academic Editor: Vasile Palade

Received: 23 December 2023

Revised: 4 February 2024

Accepted: 16 March 2024

Published: 21 March 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Traditionally, medical images are manually annotated by domain experts with special skills which makes the overall process labor intensive, expensive, slow and error-prone [1]. Automated faster and more accurate methods are critical for near real-time diagnosis

and better patient outcomes. This review focuses on the application of convolutional neural networks for medical image classifications, emphasizing recent improvements in algorithms and approaches. We covered key CNN methodologies as applied in research and clinical settings with respect to medical image localization, detection, preprocessing, segmentations and classifications. Machine learning algorithms were also applied for statistical modeling of the current literature to uncover latent topics, prevalent themes, method gaps and possible future advancements.

### *1.1. Background and Context*

Healthcare is a high priority sector due to its importance for wellness, healthspan and lifespan. Higher levels of healthcare and services entail direct, fast and reliable diagnostic approaches, such as using medical imaging. However, the interpretations of medical images by medical experts are quite limited due to subjectivity of experts and the complexity of the images [1]. In addition, extensive variations exist among experts partly attributable to human fatigue as a result of the heavy workloads of medical professionals [1].

Following the success of CNN in image processing in other real-world applications, it is also being explored as a key and robust method for applications in clinical settings [2,3]. In this review, we compiled recently improved components of deep CNN architectures, popular frameworks, activation functions, preprocessing approaches, publicly available datasets, ensemble methods and optimization techniques that are being applied for medical image understanding. Additionally, we used machine learning-based statistical modeling of the current literature to identify patterns, trends, method gaps and future advancements that were not obvious from the individual studies. This review ends with a discussion on methodological challenges and open research issues with regard to applications of CNNs for medical imaging.

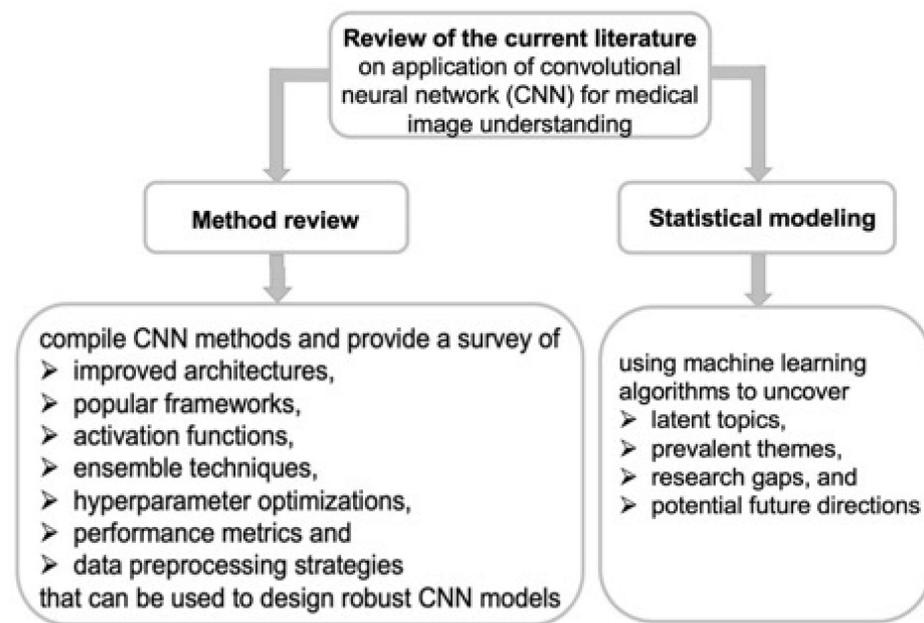
### *1.2. Importance of CNN for Medical Image Classification*

Imaging techniques are used to capture anomalies or pathological parts of the human body [4]. The captured images must be understood for the diagnosis, prognosis and treatment planning of the anomalies [4]. Analyzing images generated in clinical practice by extracting information in an efficient manner is critical for improved clinical diagnosis. However, the effectiveness of image understanding performed by skilled medical professionals is limited (and the process is slow and error-prone) due to the scarce availability of human experts and the fatigue and rough estimate procedures involved. CNNs are being widely accepted and practiced as effective tools for image understanding due to their ability to learn and extract features automatically.

There is a growing interest among researchers and clinicians in applying CNN methods for segmentation, abnormality detection, disease classification and diagnosis [5–8]. Different variations of CNN methods use different approaches to improve their performance in wide ranges of image classification tasks [9,10]. The robustness and automatability of CNNs in addition to reports of CNN techniques outperforming human experts seem to be the driving forces for the enthusiastic surge of their use in medical image understanding [11–13].

### *1.3. Objectives of the Study*

This study is designed to organize and present CNN algorithms (including improved architectures), activation functions, popular frame works, optimization approaches, relevant datasets, data preprocessing techniques and model ensemble methods in one place and make them available for researchers and clinicians who are interested in setting up their own CNN experiments (Figure 1).



**Figure 1.** Application of CNN methods for medical image understanding, and machine learning-assisted statistical modeling of the current literature.

Another important objective of this study is to use machine learning algorithms for the statistical modeling of the literature (to identify current practices and future trends that are relevant to CNN application for medical image understanding) (Figure 1).

#### 1.4. What Distinguishes the Current Study from Previously Published Review Papers?

There are great reviews focusing on the application of deep learning and CNN for medical image understanding. Many of these reviews focus on findings of the individual papers (result reviews), on specific method, on specific medical imaging or on a particular disease. But this review is both an image type and disease agnostic method review, and at the same time, it includes the machine learning-assisted statistical modeling of the current literature on the application of CNN for medical image understanding. Indeed, there are other reviews which have surveyed broader areas of the literature. And yet, the coverage of alternative CNN components in previously published review papers are less comprehensive than what is compiled in this study.

## 2. Review of CNN Algorithms and Methods

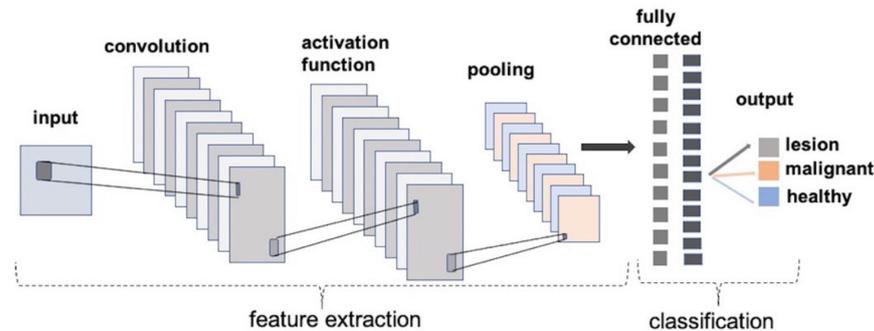
In this review, we compiled and summarized the recent advances in CNN-based medical image understanding and highlighted methodological challenges and opportunities. We began by providing an overview of the key components of CNN architectures, design improvements and activation function that are used for medical image understanding. Then, we discussed popular frameworks, ensemble techniques, widely used hyperparameters, optimizations and tuning approaches, performance metrics, databases of relevant medical images and input data preprocessing that are essential for developing more robust and transferable models. This study attempted to provide a comprehensive overview of the current state of CNN-based methods as applied for medical image understanding. Additionally, statistical modeling was used to identify some of the open CNN-related method gaps and to suggest potential future advancements.

### 2.1. Basic Architectures of CNNs

CNNs have become the mainstream algorithms for image classification due to their remarkable performance for object detection, action recognition, image classification, segmentation and disease diagnosis [7,14–22]. CNNs have the advantage of being able to

distinguish complex shapes of images [23] due to their ability to learn and extract features without the need for prior knowledge or human intervention [24].

Architectures of CNNs are designed to automatically learn and extract features from images through a series of convolutional, pooling and fully connected layers (Figure 2) [10,16].



**Figure 2.** Basic architecture of convolutional neural network showing the main components and steps involved in medical image classifications.

Each layer, in a typical CNN architecture (Figure 2), has a specific purpose:

- The input layer is the first layer which receives the input image.
- The convolutional layer is the core layer of the CNN architecture, where the convolution operation is performed using a set of learnable kernels or filters to detect edges, corners and textures (to extract features from the input data) [25]. Feature extraction may involve strides and paddings along with kernels (1).

$$W_o = \frac{W_i - F + 2P}{S} + 1 \quad (1)$$

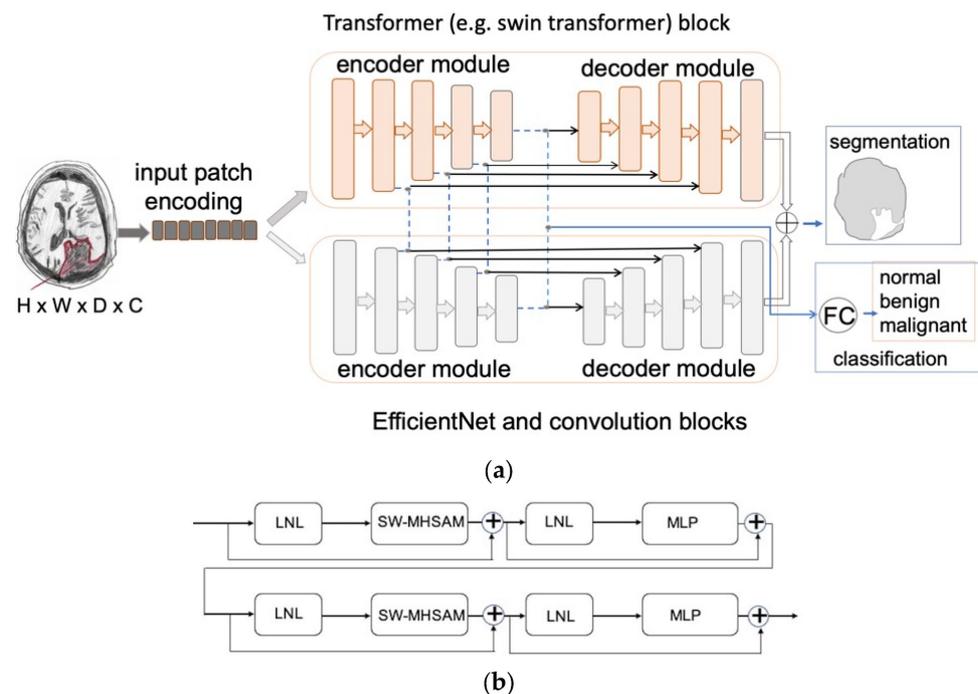
where  $W_o$  is the output size,  $W_i$  is the input size,  $F$  is the kernel size,  $P$  is the padding size and  $S$  is the size of the stride.

- The activation function introduces non-linearity to capture complex relationships in the data, and it is applied element-wise to the output of the convolutional layer.
- The pooling layer is applied to reduce the spatial dimensions (width and height) of the feature maps obtained from the convolution layer [16] by performing down-sampling.
- The fully connected layer is used to learn high-level representations by combining features learned from the previous layers.
- The output layer is the last layer which produces the desired output based on the task at hand.

CNN architectures can have additional components like dropout and normalization layers, depending on the specific application and network design.

## 2.2. Improvements in Architectural Designs of CNNs

Improved or hybrid structures of CNNs with other algorithms such as transformers (Figure 3a), recurrent neural networks (RNNs), generative adversarial networks (GANs) and shallow methods have shown better performances in medical image segmentations and classifications [26].



**Figure 3.** (a) Hybrid structure of convolutional neural networks with transformer (applied for diagnosis of brain tumor). (b) Consecutive Swin transformer block. Copyright notice: (a,b) were redrawn from Figures 1 and 2 from Wang et al. [27] (published with open access under the Creative Commons License).

Each Swin transformer block consists of residual connection, and two-layer perceptron (MLP) with Gaussian error linear units, a LayerNorm layer (LNL) and a multi-head self-attention module (Figure 3b).

The query (Q), key (K) and value (V) of the self-attention can be given as (2):

$$\text{attention}(Q, K, V) = \text{SoftMax}\left(\frac{QK^T}{\sqrt{Q_d}} + B\right)V \tag{2}$$

where  $Q_d$  is the query dimension and B is the bias matrix

High-precision medical image segmentation is a challenging task due to the existence of inherent distortion and magnification in medical images as well as the presence of lesions with similar density to normal tissues. Recently, hybrid structures of transformers with CNN along with attention blocks have been designed and progressively improved to tackle the problem of medical segmentation. Many of the hybrid and/or improved structures are also designed for medical image classifications in addition to segmentation tasks (Table 1).

**Table 1.** Improved or hybrid CNN architectures that are applied for medical image understanding.

CNN Design	Description	Specific Application
U-net [28] U-net++ [29]	U-shaped network design or a nested U-net architecture.	For the segmentation of medical images [30–32].
attention U-net [33]	Attention gate (AG) model.	Automatically learns to focus on structures of varying sizes and shapes.
ResNet [34–36]	A deep residual learning network (a shortcut connection model to significantly reduce the difficulty of training very deep CNNs).	Aims to simplify very deep networks by introducing a residual block that sums two input signals.

Table 1. Cont.

CNN Design	Description	Specific Application
FC-DenseNet [37,38]	Fully convolutional DenseNet developed by the composition of dense blocks and pooling operations in which the up-sampling path was introduced to restore the input resolution.	For semantic image segmentation.
ViT [39,40]	Vision transformer.	For image segmentation.
Swin Transformer [41]	Hierarchical vision transformer using shifted windows (uses a sliding window to limit self-attention calculations to non-overlapping partial windows).	Serves as a general-purpose backbone for medical image segmentation and classification.
Swin UNETR [42] UNETR [43]	Shifted windows UNet transformers (Swin UNet Transformers): pretrained, large-scale and self-supervised 3D models for data annotation (tailored for 3D segmentation and directly use volumetric data).	Pretrained framework tailored for self-supervised tasks in 3D medical image analysis.
ResViT [44]	Residual vision transformers.	Generative adversarial network for multi-modal medical image synthesis.
TransUNet [45]	This has an embedded transformers in the down-sampling process to extract the information in the original image.	To solve a lack of high-level detail.
TFCNs [46]	Transformers for fully convolutional DenseNet.	To tackle the problem of high-precision medical image segmentation by introducing a ResLinear-transformer and convolutional linear attention block to FC-DenseNet
SETR [47]	Segmentation transformer.	A pure transformer (without convolution and resolution reduction) to encode an image as a sequence of patches.
Deformable DETR [48]	Fully end-to-end object detector using a simple architecture by combining CNNs and transformer encoder–decoders architecture.	Mitigates the slow convergence and high complexity issues of DETR.
Medical Transformer [49]	Gated axial attention for medical image segmentation.	Operates on the whole image and patches to learn global and local features.
O-Net [27]	Framework with deep fusion of CNN and transformer.	For simultaneous segmentation and classification.
TransMed [50]	Combines CNN and transformer to efficiently extract low-level features of images.	Multi-modal medical image classification.
SMESwin Unet [51]	Superpixel and MCCT-based channel-wise cross-fusion transformer (CCT) coupled with multi-scale semantic features and attention maps (Swin UNet).	For medical image segmentation.
CNN-SVM hybrid [52]	Threshold segmentation approach.	For tumor detection and classification of MRI brain images.
SegTransVAE [53]	Hybrid CNN–transformer with regularization.	For medical image segmentation.
autoencoder-hybrid CNN-LSTM model [54]	Hybrid of CNN with RNN.	For COVID-19 severity prediction from lung ultrasounds.
HCT-Net [55]	Hybrid CNN–transformer model based on a neural architecture search network.	A neural architecture search network for medical image segmentation.
HybridCTrm [56]	Bridging CNNs and transformers.	For multimodal image segmentation.

Table 1. Cont.

CNN Design	Description	Specific Application
MFDNN [57]	MFDNN (multimodal fusion deep neural network) integrates different modalities (medical imaging, genomics, clinical data) to enhance lung cancer diagnostic accuracy.	Used for lung cancer classification by integrating clinical data, electronic health records and multimodal approaches (to improve the accuracy and reliability of lung cancer diagnosis).
M <sup>3</sup> BTCNet [58]	This architecture uses metaheuristic deep neural network features optimization.	For a multi model brain tumor classification.
NeuroNet19 [59]	Uses VGG19 as its backbone and incorporates inverted pyramid pooling module (iPPM) to capture multi-scale feature maps (to extract both local and global image contexts). Local interpretable model-agnostic explanations (LIME), is used to highlight the features or areas focused on while predicting individual images.	An explainable deep neural network model for the classification of brain tumors (using MRI data).
sparse deep neural network features [60]	Designed based on dense neural network (VGG-16 and Resnet-50) and sparse neural networks (inception v3).	For detection and classification of non-small cell lung cancer types.
3D CNN Multimodal Framework [61]	The framework comprises a 3D CNN for each modality, whose predictions are then combined using a late fusion strategy based on Dempster–Shafer theory.	Classification of MRI images with multimodal fusion. This multimodal framework processes all the available MRI data in order to reach a diagnosis.
Feature fusion-based ensemble CNN learning optimization [62]	An ensemble CNN framework incorporates optimal feature fusion: multiple CNN models with different architectures are trained on the dataset using fine-tuning and transfer learning techniques.	For the automated detection of pneumonia. Learning optimizations achieved by iteratively eliminating irrelevant features from the fully connected layer of each CNN model using chi-square and mRMR methods. Optimal feature sets are then concatenated to enhance feature vector diversity for classification.
External validation of a deep learning model [63]	This model is based on ResNet-18 to automatically assess the mammographic breast density (for each mammogram), providing a quantitative measure of the breast tissue composition.	For breast density classification.
Weakly Supervised Deep Multiple Instance Learning [64]	This is a two-stage framework based on deep multiple instance learning. It requires only global labels (weak supervision).	For diagnosis and detection of breast cancer. This approach provides classification of the whole volume and of each slice and the 3D localization of lesions through heatmaps.
DM-CNN [65]	Dynamic multi-scale CNN contains four sub-modules: dynamic multi-scale feature fusion module (DMFF), hierarchical dynamic uncertainty quantifies attention (HDUQ-Attention), multi-scale fusion pooling method (MF Pooling) and multi-objective loss (MO loss)	For medical image classification with uncertainty quantification. DMFF selects convolution kernels according to the feature maps of each level for information fusion. HDUQ-Attention has a tuning block to adjust the attention weight according to the information of each layer. The Monte Carlo (MC) dropout structure is for quantifying uncertainty. The MF pooling is to speed up the computation and prevent overfitting. And the MO loss is for a fast optimization speed and good classification effect.

Table 1. Cont.

CNN Design	Description	Specific Application
lightweight multi-modality UNeXt and Wave-MLP semantic segmentation network [66]	The wave block module in Wave-MLP replaces the Tok-MLP module in UNeXt. The phase term in wave block can dynamically aggregate token to improve segmentation accuracy. An AG attention gate module at the skip connection suppresses irrelevant feature representations. Then, the focal Tversky loss is added to handle both binary and multiple classification tasks.	For multi-modality medical image semantic segmentation.
MobileNets [67]	Efficient CNN based on a streamlined architecture that uses depth-wise separable convolutions to build light-weight deep neural networks.	For mobile and embedded vision applications. Use cases include object detection, fine-grain classification, face attributes and largescale geo-localization.
UL-BTD [68]	an automated ultra-light brain tumor detection (UL-BTD) system based on ultra-light deep learning architecture (UL-DLA) for deep features, integrated with highly distinctive textural features, extracted by gray level co-occurrence matrix.	For multiclass brain tumor detection. It forms a hybrid feature space for tumor detection using support vector machine, leading to high prediction accuracy and optimum false negatives with limited network size to fit within the average GPU resources of a modern PC system.

2.3. Activation Functions Used in CNNs

The use of an optimal activation function along with a robust CNN structure is important for medical image analysis. Having a suitable nonlinear activation function can significantly improve the performance of the network. It is important to note that there is no single or “best” activation function that works universally for all CNN architectures. The choice of activation function should be based on empirical evaluation and specific task requirements (Table 2).

Table 2. Activation functions frequently used in CNN applications for medical image processing. The middle solid line separates the ReLU families and derivates from other classes of activation functions.

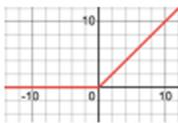
Activation Function	Equation	Graphical Representation *	Short Description
Rectified Linear Unit (ReLU)	ReLU(x) = max(0, x) i.e., ReLU(x) = {0, if x ≤ 0, x, if x > 0}		Computationally efficient and helps alleviate the vanishing gradient problem, allowing for faster training and improved network performance.
Leaky ReLU	Leaky ReLU(x) = max(alpha × x, x) = {x, if x > 0, alpha × x, if x ≤ 0} x is the input, and alpha is a small positive constant (determines the slope for negative input values).	 alpha = 0.1	Addresses the issue of “dead neurons” by allowing small negative values instead of setting them to zero; provides some gradient flow for negative inputs during backpropagation.
Parametric ReLU (PReLU)	PReLU(x) = {x, if x > 0, alpha × x, if x ≤ 0}; Alpha is a parameter that can be learned during the training process (controls the slope for negative input values).	Similar to leaky ReLU (although alpha here is a parameter to be learned and optimized).	During training, the alpha parameter is updated through backpropagation, enabling the network to learn the optimal value for each neuron. Adjusting the slope for negative inputs can lead to improved performance and better representation learning.

Table 2. Cont.

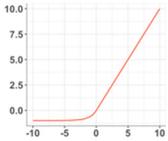
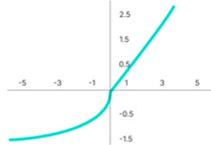
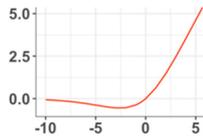
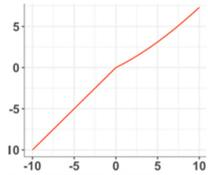
Activation Function	Equation	Graphical Representation *	Short Description
Randomized Leaky ReLU (RReLU)	The slope is fixed to a predefined value during testing. This introduces a form of regularization and can help prevent overfitting.	Similar to leaky ReLU.	A variation of Leaky ReLU that randomly samples the slope from a uniform distribution during training.
Exponential Linear Unit (ELU)	$ELU(x) = \{x, \text{ if } x > 0, \\ \alpha \times (\exp(x) - 1), \text{ if } x \leq 0\}$ Alpha is a hyperparameter (controls the behavior of the function); ELU captures more nuanced information from negative inputs and alleviate the vanishing gradient problem.	 <p>alpha = 1.0</p>	Smooths negative inputs by using an exponential function; the exponential smoothing helps reduce the impact of noisy activations.
Scaled Exponential Linear Unit (SELU)	$SELU = \lambda \begin{cases} a(e^x - 1) & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$ with predefined value for lambda $\lambda$ or in general $SELU(x) = \{scale \times (x \text{ if } x > 0 \text{ else } (\alpha \times \exp(x) - \alpha))\}$ , if training; $scale \times x$ , if testing} x is input to the activation function, alpha is a hyperparameter that controls the slope for negative inputs and scale is a scaling factor to maintain the mean and variance of the inputs close to 0 and 1, respectively. SELU has the property of self-normalization, which can lead to improved performance and stability in deep neural networks.	 <p>SELU, by adjusting the mean and variance, takes care of internal normalization. Gradients can be used to adjust the variance (needs a region with a gradient &gt; 1 to increase it).</p>	During training, SELU applies a modified ELU function (negative inputs are transformed with a negative slope). The scale factor stabilizes the activations and ensures self-normalization. The mean and standard deviation of the outputs are enforced to be approximately 0 and 1, respectively (helps address the vanishing/exploding gradient problem). During testing, SELU behaves as a scaled identity function (inputs are multiplied by the scale factor to preserve the output magnitude).
Swish	$SWISH(x) = x \times \text{sigmoid}(\beta \times x)$ Beta is a hyperparameter that controls the behavior of the function. Higher values of beta can lead to more pronounced non-linearity, while lower values can make it closer to the identity function.	 <p>beta = 0.5</p>	Combines the linearity of the identity function (x) with the non-linearity of the sigmoid function (for positive inputs: retains the linearity; for negative inputs, the output towards zero is dampened due to the sigmoid function). It performs well in CNNs.
SWISH-RELU	$SWISH-RELU(x) = x \times \text{sigmoid}(\beta \times x) \text{ if } x > 0 \\ SWISH-RELU(x) = x \text{ if } x \leq 0$ The advantage of SWISH-RELU is that it retains the desirable properties of Swish, such as the smoothness and non-monotonic behavior, while also providing a fallback to ReLU for negative inputs. This fallback mitigates the problem of dead neurons and vanishing gradients associated with the standard Swish activation function.	 <p>beta = 0.1</p>	The Swish activation function with a ReLU fallback is a Swish and ReLU hybrid. The sigmoid introduces a smooth non-linearity, while the ReLU fallback ensures that the activation does not completely vanish for negative inputs. SWISH-RELU performs well in CNNs for image classification.

Table 2. Cont.

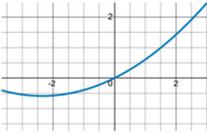
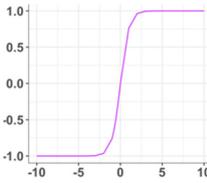
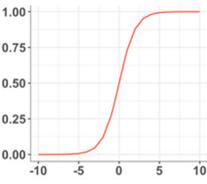
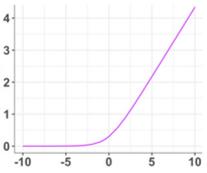
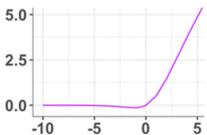
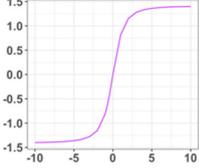
Activation Function	Equation	Graphical Representation *	Short Description
Gaussian Error Linear Unit (GELU)	$\text{GELU}(x) = 0.5x \times (1 + \text{erf}(x/\sqrt{2}))$ <p>This is smooth and non-monotonic. <math>x</math> is the input and erf is the error function used to model cumulative distribution.</p>	 <p>erf = 0.3</p>	GELU has a smooth and non-linear behavior that can help capture complex patterns and gradients; it performs well in NLP and CNNs. It is computationally more expensive than ReLU due to the involvement of erf but improves the performance in certain scenarios.
Softmax	$\text{Softmax}(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)}$ <p>Given an input vector of <math>x = [x_1, x_2, \dots, x_n]</math>, the Softmax function computes the probability <math>p_i</math> for each element <math>x_i</math> as:  <math display="block">\text{Softmax}(x_j) = \exp(x_j) / \sum(\exp(x_j))</math> for <math>j = 1</math> to <math>n</math>  The highest probability class is selected as the predicted class label.</p>	<p>Boundaries vary based on the <math>x_i</math> and <math>x_j</math> values.</p>	Used as the final activation function in the output layer for multi-class classification tasks (takes a vector of real numbers inputs and outputs a vector of probabilities between 0 and 1 that sum up to 1). Enables the network to assign probabilities to each class, indicating the model's confidence for each class prediction.
Hyperbolic Tangent (Tanh)	$\text{Tanh}(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}$ $\exp(-x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ <p>Non-linear symmetric function around the origin (squeezes the input value into a range between <math>-1</math> and <math>1</math>).</p>		Useful for tasks that require outputs in the range of $-1$ to $1$ or for modeling symmetric patterns. Suffers from the "vanishing gradient" problem, where the gradient becomes extremely small for inputs with very high absolute values.
Sigmoid (logistic)	$\text{sigmoid}(x) = 1 / (1 + \exp(-x))$ <p>A non-linear function that squeezes the input value into a range between 0 and 1. Suffers from the "vanishing gradient" problem, where the gradient becomes extremely small for inputs with very high or very low absolute values.</p>		Maps any real-valued number to a value between 0 and 1, with values close to 0 representing the lower end of the range and values close to 1 representing the upper end (suitable for binary classification tasks or probabilistic outputs).
Softplus	$\text{Softplus}(x) = \log(1 + \exp(x))$ <p>Designed to be a smooth and differentiable approximation of the ReLU function, which is non-differentiable at <math>x = 0</math>. Commonly used in variational autoencoders (VAEs) and some recurrent neural networks (RNNs).</p>		Has similar properties to ReLU, where positive inputs are passed through unchanged, while negative inputs are mapped to small positive values. It introduces non-linearity to the network, allows for the modeling of complex patterns, and provides smoother gradients than ReLU (facilitates better training and convergence).
Mish	$\text{Mish}(x) = x \times \tanh(\text{softplus}(x))$ <p>Mish does not have a closed-form derivative and is often approximated or numerically computed during backpropagation. It introduces non-linear behavior, captures complex patterns and alleviates the vanishing gradient problem.</p>	 <p>Performs well in image classification and NLP.</p>	Mish combines the non-linearity of the softplus function with the smoothness of the hyperbolic tangent function (has a similar shape to the Swish activation function but with a gentler slope for negative inputs).

Table 2. Cont.

Activation Function	Equation	Graphical Representation *	Short Description
Inverse Square Root Unit (ISRU)	$ISRU(x, \alpha) = x / \sqrt{1 + \alpha \times x^2}$ <p>Alpha is a positive constant that determines the steepness and shape of the ISRU function; a larger alpha value results in a steeper curve, while a smaller alpha value leads to a more gradual curve. The square root and normalization ensure that the output remains within a reasonable range.</p>	 <p>alpha = 0.5</p>	ISRU is used as an alternative to sigmoid or tanh in situations where a more gradual transition from low to high activations is desired. But it is not widely used in deep learning models.

\* Graphs of activation functions were generated using ggplot2 package of R or an online graphing method called desmos (<https://www.desmos.com>, accessed: 14 June 2023). Many of the activation functions (Table 2) are smooth and non-monotonic, meaning that they do not strictly increase or decrease for all input values.

Other activation functions that are not commonly used, in comparison to the ones listed in Table 2, include the ReLU family (Symmetric ReLU (SReLU), inverse square root linear unit (ISRLU), leaky ReLU with arbitrary slope (LReLU), randomized leaky ReLU (RReLU)), continuously differentiable exponential linear unit (CELU), Softsign, Maxout, squeeze-and-excitation nonlinearity (SQNL), sine, cosine, ArcTan, hard sigmoid, hard tanh, linearly scaled hyperbolic tangent (LiSHT), bent identity, the bent identity family (bent identity parameterized, bent identity smooth, bent identity parametric) and the Ada family (AdaBound, AdaBelief, AdaM, AdaMod, AdaShift, AdaSign, AdaBoundW, AdaBeliefW, AdaMW, AdaModW, AdaShiftW, AdaSignW, AdaBoundT, AdaBeliefT, AdaMT, AdaModT and AdaShiftT).

#### 2.4. Popular Frameworks

There are a number of CNN frameworks widely used for medical image understanding (Table 3). The popularity of frameworks is ranked based on the number of search-hits using Google Scholar, PubMed and IEEE Xplore (Figure 4 and Table 3). The use of Keras as a TensorFlow interface seems to be the most widely used framework across the three search engines/databases (Figure 4).

Table 3. Frequently used CNN frameworks (the order of this list is arbitrary).

Framework * (Repository)	Developed (Maintained)	Short Description	Numbers of Hits for Each Search Engine		
			Google Scholar	PubMed	IEEE Xplore
TensorFlow	Google Brain Team	An end-to-end machine learning platform.	231,000	364	2154
PyTorch	Meta AI	Based on the Torch library.	82,600	196	650
Theano	Montreal Institute for Learning Algorithms	Allows the definition, optimization, and efficient evaluation of mathematical expressions involving multi-dimensional arrays.	29,600	23	69
Keras	François Chollet	Provides a Python interface for ANNs, e.g., TensorFlow.	697,000	182	893
MXNet	Apache Software Foundation	Scalable, allows fast model training and supports multiple programming languages.	7930	5	59
Caffe/Caffe2	University of California, Berkeley	A lightweight, modular and scalable deep learning framework.	7410	No results	No result

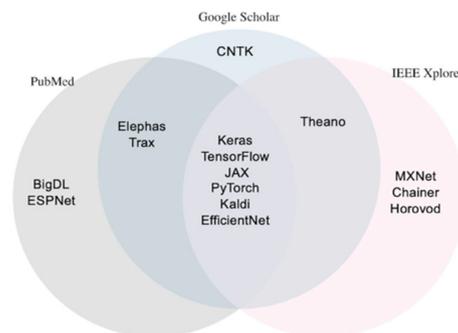
Table 3. Cont.

Framework * (Repository)	Developed (Maintained)	Short Description	Numbers of Hits for Each Search Engine		
			Google Scholar	PubMed	IEEE Xplore
Chainer	Preferred Networks, Inc., Tokyo, Japan	A collection of tools to train and run neural networks for computer vision tasks.	5700	9	52
CNTK	Microsoft	Describes neural networks as a series of computational steps via a directed graph.	19,800	8	28
Torchnet	PyThorch TNT	An abstraction to train neural networks (for logging and visualizing, loading and training).	97	61	No results
JAX	Google	Provides interfaces to compute convolutions across data.	143,000	2692	113
EfficientNet	ICML 2019: Mingxing Tan, Quoc V. Le	A CNN architecture and scaling method that uniformly scales all dimensions of depth/width/resolution using a compound coefficient.	23,700	325	648
SRnet	Niwhskał/SRNet	A twin discriminator GAN that can edit text in any image while maintaining context of the background, font style and color.	2960	9	34
LFnet	Learning Local Features from Images	Deep architecture to learn local features and can be trained end-to-end with just a collection of images, from scratch, without hand-crafted priors.	1150	5	6
Horovod	The Linux Foundation	Distributed deep learning training framework for TensorFlow, Keras, PyTorch and Apache MXNet.	1840	No results	40
Attention Factorization Machine (AFM)	Jun Xiao et al. [69]	Learning the weight of feature interactions via attention networks.	18,500	No results	3
Neural Factorization Machine NFM-PT	Xiangnan He and Tat-Seng Chua	For sparse predictive analytics (or prediction under sparse settings).	4480	2	1
Deep Factorization Machine (DeepFM)	Guo, H et al. [70]	Combines the power of factorization machines for recommendation and deep learning for feature learning with no need of feature engineering besides raw features.	2340	3	20
Deep Cross-Network (DCN)	Wang, R et al. [71]	Applies feature crossing networks at each layer that do not require manual feature engineering, and hence, it is more efficient in learning certain bounded-degree feature interactions.	581	1	4
Trax	Google Brain Team	An end-to-end library for deep learning that focuses on clear code and speed.	20,800	156	38
Kaldi	DNN in KALDI	An open-source speech recognition toolkit.	47,100	73	174
OpenSeq2Seq	FazedAI/ OpenSeq2Seq	A TensorFlow-based toolkit for sequence-to-sequence models.	127	No results	No results

Table 3. Cont.

Framework * (Repository)	Developed (Maintained)	Short Description	Numbers of Hits for Each Search Engine		
			Google Scholar	PubMed	IEEE Xplore
ESPNet/ ESPNet	Watanabe, S et al. [72]	An end-to-end toolkit for speech processing, recognition and text-to-speech translation.	3270	140	30
wav2letter++	Pratap, V et al. [73]	A fast open-source deep learning speech recognition framework.	4	No results	3
Elephas	Max Pumperla and Daniel Cahall	An extension of Keras, which allows the running of distributed deep learning models at scale with Spark.	52,700	786	3
Tfaip/tfaip	Python community	Research framework for developing, organizing, and deploying deep learning models powered by TensorFlow.	26	5	No results
BigDL	Dai, J et al. [74]	A distributed deep learning library for Apache Spark (fast, distributed and secure AI for big data).	455	1684	10

\* Additional frameworks implementing graph neural networks (GNN) are available, such as PyTorch Geometric (PyTorch), TensorFlow GNN (TensorFlow) and jraph (Google JAX). Relevant application domains for GNNs include natural language processing, social networks, citation networks, molecular biology, physics and NP-hard combinatorial optimization problems.



**Figure 4.** Overlap among the top 10 most popular CNN frameworks that were ranked based on the search hits using Google Scholar, PubMed and IEEE Xplore.

### 2.5. Ensemble Approaches for CNN Models

There are several models that can be ensembled with CNN designs which can be used for medical image analyses. Ensemble techniques aim to improve the robustness and accuracy of CNNs [75–77]. Ensemble methods for CNN models include mixture ensemble of CNNs [78] used for breast tumor classification [79], ensembles of pre-trained CNNs (such as inception v3) [80] used for epilepsy classification [78], in-network ensembles for obstructive sleep apnea detection [81,82], weighted average ensembles for pneumonia detection [83], a self-ensemble framework [84] used for brain lesion segmentation, orthogonal and attentive ensemble networks [85] used for COVID-19 diagnosis [86,87], 3D CNN ensembles used for pulmonary nodule classification in lung cancer screening [88] and ensembles of REFINED-CNN built under different choices of distance metrics and/or projection schemes used for anti-cancer drug sensitivity prediction [89]. Ensembled designs consisting of deep CNN and recurrent neural network architecture are applied for the recognition of end-to-end arousal from ECG signals [90].

### 2.6. Hyperparameters of CNNs Used for Medical Image Analyses

Hyperparameters are settings that affect the performance of Neural nets. There are more than 420 hyperparameters reported in the literature used for tuning up deep neural

networks (Supplementary Table S1). Of the 420 plus, around 30 to 40 hyperparameters are widely reported in relation with application of CNN for object recognition and medical image understanding (Table 4).

**Table 4.** The most widely used hyperparameters for convolutional neural networks (CNNs).

Hyperparameter *	Description
Learning rate	Controls the step size at each iteration during training and influences how quickly the model learns.
Number of epochs	Determines the number of times the entire dataset is passed through the network during training.
Batch size	Specifies the number of training examples in each mini-batch used for updating the model's parameters.
Batch normalization	A normalization technique that helps stabilize the learning process by normalizing the inputs of each layer.
Optimizer type	Selects the optimization algorithm used to update the model's weights based on the computed gradients (e.g., Adam, SGD).
Loss function	Defines the objective function used to measure the difference between predicted and actual values (e.g., categorical cross-entropy, mean squared error).
Activation function	Applies non-linearity to the output of a neuron and determines the range of values that can be produced by the layer (e.g., ReLU, sigmoid).
Dropout rate	Controls the probability of randomly setting a fraction of the input units to 0 during training, reducing overfitting.
Weight initialization strategy	Determines how the initial weights of the model are set before training begins.
Number of layers	Specifies the depth or the number of layers in the CNN architecture.
Filter/kernel size	Defines the spatial extent of the filters (convolutional kernels) used to scan the input data.
Pooling type	Determines the downsampling operation applied to reduce the spatial dimensions of the feature maps (e.g., max pooling, average pooling).
Pooling size	Specifies the size of the pooling window used for downsampling.
Stride	Defines the step size at which the filter/kernel moves horizontally or vertically when performing convolutions or pooling.
Padding	Determines whether and how extra border pixels are added to the input data before performing convolutions or pooling.
Learning rate decay	Reduces the learning rate over time to allow for finer adjustments during training.
Weight decay	Adds a penalty term to the loss function to discourage large weights, reducing overfitting.
Data augmentation	Applies random transformations to the training data, such as rotation, flipping or zooming, to increase the diversity of examples and improve generalization.
Transfer learning	Uses pre-trained models on large-scale datasets as a starting point for training on a specific task, saving training time and potentially improving performance.
Early stopping	Stops the training process if the validation loss does not improve over a certain number of epochs, preventing overfitting and saving computational resources.
Learning rate schedule	Specifies how the learning rate is adjusted during training, such as by reducing it after a certain number of epochs or based on a predefined schedule.
Initialization of biases	Determines how the biases of the model's layers are initialized.
Learning rate warm-up	Gradually increases the learning rate at the beginning of training to stabilize the optimization process.
Image normalization	Specifies how the input images are normalized (e.g., mean subtraction, scaling to a certain range).
Network architecture	Defines the overall structure of the CNN model, including the arrangement and types of layers (e.g., VGG, ResNet, Inception).
Number of filters per layer	Determines the depth of the feature maps produced by each convolutional layer.

Table 4. Cont.

Hyperparameter *	Description
Dilated convolutions	Allows the network to have a larger receptive field without increasing the number of parameters.
Weight sharing	Shares weights across different parts of the network to reduce the number of parameters and improve generalization.
Learning rate annealing	Gradually decreases the learning rate during training to fine-tune the model's parameters.
Input image size	Specifies the size of the input images to the CNN model.
Number of convolutional layers	Determines the depth or capacity of the CNN; the appropriate number depends on the task, the size and diversity of the dataset and the computational resources.
Number of fully connected layers	Used to map the high-level features to the desired output. The number of neurons or units in each fully connected layer is another hyperparameter.
Momentum	Used in optimization algorithms (e.g., SGD) and can improve the convergence speed and stability of CNN training by accumulating momentum from past gradients.
Inverse dropout	Makes inference faster during test time.
L2 regularization	For the sparse representation of features.

\* These hyperparameters offer a wide range of options for configuring and optimizing CNN models for various tasks, including medical image classification. The optimal values for these hyperparameters depend on the specific task and dataset being used and can be determined through hyperparameter tuning.

### 2.6.1. Hyperparameter Tuning and Optimization Methods

The tuning of deep learning architectures helps to improve the ease of data encoding, integrative layering, multivariate classification and predictive model performance. Particularly, the hyperparameter tuning of CNN models are important steps for training, iterative tuning and benchmarking (to make classifications).

The following important and widely used CNN hyperparameter tuning methods can be used to improve the reproducibility of model outputs or performances:

- Automatic hyperparameter optimization tools (like Amazon's HyperparameterTuner or Google Vizier) [91,92]
- Optimization algorithms, such as particle swarm [93], black-box and gradient or Bayesian-based algorithms (such as the surrogate-based [94,95] or asymmetric kernel function [96]) or genetic or custom genetic algorithms [97], artificial bee colony algorithm [98], the firefly algorithm [99] and the Broyden–Fletcher–Goldfarb–Shanno algorithm (for iteratively solving unconstrained nonlinear optimization)
- Search methods, such as grid search [100,101], random search, random grid coarse-to-fine search [102], weighted random search [102], tabu search and elastic search/BM25 similarity search.
- Metaheuristic optimization techniques [103] or SHO metaheuristic optimization for fine-tuning the weights, biases and hyperparameters [104]
- The orthogonal array tuning method [2], the adaptive hyperparameter tuning and the covariance matrix adaptation evolution strategy.
- Simulated annealing, the KNN approach, per-parameter regularization and the EVO technique (used to obtain the accurate optimized value in terms of hybridized exploitation and exploration).

Some tuning methods may be more computationally expensive than others, so it is important to consider the trade-offs between accuracy and efficiency when selecting a hyperparameter tuning method.

### 2.6.2. Tuning of Parameters

CNN model parameters such as weights and biases can be randomly initialized and iteratively updated (using backpropagations) guided by the markers of model performance (described under model performance). Applying factorizations to the weight metrics in

the networks can help to significantly reduce the total number of parameters to be trained. Different gradient descents, including the stochastic gradient descent and exponentiated gradient algorithms can be used to update parameters [105].

### 2.6.3. Benchmarking of Model Performances

Model performances can be benchmarked by calculating model convergence, cost, and training set and validation set errors. The quantitative values of the training and validation set errors will be evaluated in reference to the base error on datasets that are from the same distribution. If the training set error is high, the model has a high bias (underfitting) toward the training set. To address the high training set error (high model bias or underfitting), it is recommended to use deeper neural nets, longer training, and/or different CNN architecture. On the other hand, if the development set error is high while training set error is low, the model has high variance (overfitting) toward the training set. To address the high validation set error (high model variance or overfitting), it is recommended to use more datasets of the same distribution (e.g., publicly available databases), regularization, a different neural network architecture and/or inverse dropout.

#### Performance Metrics Used in Evaluating CNN Models

The performance of classifier models can be evaluated using the diagnostic (confusion) matrix and derivatives of the main diagnostic parameters: such as sensitivity (recall), specificity (true negative rate), F1-score, positive and negative predictive values, accuracy, precision, positive and negative likelihood ratios, diagnostic odds ratio, Matthew's correlation coefficient and the area under the receiver operating characteristic curve (both on the validation and test datasets).

Confusion matrix: True class (columns) and Predicted class (rows)

True Positives (TP)	False Positives (FP)
False Negatives (FN)	True Negatives (TN)

- Classification accuracy is the percentage of correctly classified instances out of the total number of instances in the dataset (3).

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (3)$$

Accuracy is used to evaluate the performance of CNNs in image classification tasks [106,107].

- Sensitivity and specificity are measures of the true positive rate and true negative rate, respectively. Sensitivity measures the proportion of correctly identified actual-positives (4), and specificity measures the proportion of correctly identified actual-negatives (5).

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (5)$$

These metrics are commonly used in medical image analysis tasks [108].

- The F1 score is a measure of the balance between precision and recall, which are metrics that evaluate the accuracy and completeness of the model's predictions, respectively. That is, the F1 score (6) is the harmonic mean of precision and recall and is used to evaluate the performance of CNNs in binary classification tasks.

$$\text{F1 score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \quad (6)$$

- Mean squared error (MSE): measures the average squared difference between the predicted and actual values (7). MSE is particularly important to evaluate quantitative or regression tasks.

$$\text{MSE} = \frac{1}{n} \sum_{k=0}^n (y_i - \hat{y}_i) \quad (7)$$

where  $n$  is the number of observations,  $y_i$  is the observed values and  $\hat{y}_i$  is the predicted values.

- Receiver operating characteristic (ROC) is a graphical representation of the trade-off between sensitivity and specificity for different classification thresholds. It is used to evaluate the performance of CNNs in binary classification tasks [2,5].
- Area under the curve (AUC) is a summary measure of the ROC curve representing the probability that a randomly chosen negative instance will be ranked lower than a randomly chosen positive instance. It is commonly used to evaluate the overall performance of CNNs in binary classification tasks.

The generalizability of classifier models can be further evaluated on totally independent datasets of similar distribution.

### 2.7. Data Pre-Processing Methods

- Data distillation methods: uniform experiment design method, highlighting, background filling, resizing, noise reduction, the Gabor filter model, image defect detection and implicit differentiation;
- Optical flow image processing;
- Sliding window data-level approach [109];
- Flattening and normalizing data in a task-specific manner;
- One-hot vector encoding method;
- Frequency-based tokenization [110];
- Training-validation-testing splits.

These pre-processing techniques are used to prepare the image dataset for CNN modeling. In addition to pre-processing, some studies also use segmentation via CNN to further analyze the images.

### 2.8. Image Datasets Relevant for Medical Themes

The use of CNN techniques in medical image analysis and disease classification necessitates the availability of comprehensive and diverse datasets [111–113]. The success of these techniques relies on the richness and representativeness of the datasets, as they enable the extraction of salient information and features from medical images and records [112,114,115].

The list of salient datasets important for medical themes (Table 5) encompasses a wide array of medical images for diverse pathological conditions. The quality, representativeness, and diversity of these datasets make them valuable for practicing and for setting up CNN experiments as well as for CNN-based medical image understanding. Supplementary Tables S3–S5 also show lists of comparisons of performances of CNN methods applied to other popular public datasets.

**Table 5.** List of salient image datasets (that are important for medical themes).

Dataset Name (Hyperlinked URL as of 21 January 2024)	Theme	Description
Medical Information Mart for Intensive Care III (MIMIC-III)	Critical care	Electronic health records (EHR) of ICU patients, including clinical notes, demographics, lab results and imaging reports
The Cancer Genome Atlas (TCGA)	Cancer genomics	Images and genomic and clinical data for various cancer types, including gene expression, epigenetic marks, mutations and clinical outcomes.
NIH Chest X-ray dataset	Chest X-ray imaging	An open dataset of chest X-ray images labeled for common thoracic diseases, including pneumonia and lung cancer, often used for developing and evaluating image classification models.
Alzheimer’s Disease Neuroimaging Initiative (ADNI)	Neuroimaging (Alzheimer’s disease)	Longitudinal MRI and PET imaging data for Alzheimer’s disease research.
Diabetic Retinopathy Detection	Ophthalmology	Fundus images for diabetic retinopathy classification, used to develop algorithms for automated disease detection.
PhysioNet Challenge	Various cardiology and physiological signals	Datasets from PhysioNet challenges cover a variety of themes, including heart rate, blood pressure and electrocardiogram (ECG) signals.
Multimodal Brain Tumor Segmentation Challenge (BraTS)	Neuroimaging (brain tumor)	MRI images for brain tumor segmentation, challenging researchers to develop algorithms for tumor detection and segmentation.
UCI Machine Learning Repository-Health Datasets	Various	A collection of health-related datasets covering different topics, including diabetes, heart disease and liver disorders.
PhysioNet/MIMIC-CXR Database	Chest X-ray imaging	A dataset of chest X-ray images with associated radiology reports, supporting research in chest radiography.
Skin Cancer Classification-Refugee Initiative (SCC-RI)	Dermatology	A dataset of skin lesion images for skin cancer classification, focusing on refugee populations.
Federal interagency traumatic brain injury research (FITBIR)	Traumatic brain injury	These datasets include images on traumatic brain injury patients, clinical and molecular datasets.

### 2.9. Data Augmentation for Training a Robust CNN Diagnostic Model for Cases with Insufficient Training Data

Data augmentation is a critical component in training robust convolutional neural network (CNN) models when there are no sufficient training datasets. It involves generating additional data to enhance the training process, improve model performance and generalization. Augmenting the training dataset involves applying various transformations to the original images, creating new variations such as rotation, flipping, zooming, translation, brightness and contrast adjustment, Gaussian noise (adding a small amount of Gaussian noise to the images to make the model more robust to noise), elastic deformations (applying elastic deformations to the images to introduce distortions, making the model more tolerant to deformations in the input data), color jittering (randomly change the hue, saturation and brightness of the images to introduce variations in color), random cropping (a portion of the image, forcing the model to focus on different regions) and shearing (apply shearing transformations to the images, simulating changes in the viewing angle) [116–118]. These processes help the model become more robust by exposing it to different perspectives, orientations and conditions. Studies have demonstrated that data augmentation, when combined with fine-tuning and transfer learning, can significantly enhance model accuracy [119,120]. Additionally, data augmentation can be used to enhance the robustness of CNN models to noise for improved training [121,122] and to mitigate kernel saturation (to increase classification accuracy) [123]. Therefore, data augmentation

techniques can be used to develop robust CNN diagnostic models by addressing limited training data, noise and kernel saturation.

Data augmentation can be implemented using libraries such as TensorFlow's ImageDataGenerator or PyTorch's transforms. These libraries provide convenient tools for applying various transformations to the training data on-the-fly during the model training process. When implementing data augmentation, it is essential to strike a balance. For example, too much augmentation may result in the model memorizing augmented images rather than learning useful features. Additionally, it is important to consider the nature of the diagnostic task; for medical imaging, it is advisable to be cautious with certain transformations to avoid introducing unrealistic artifacts.

#### Enhancing CNN-Based Image Classification for Rare Diseases through Data Augmentation

The scarcity of labeled data for images associated with rare diseases poses a significant challenge for training accurate and robust models. The key challenges include (i) the limited availability of annotated images not only hinders the training of CNNs but also poses a risk of overfitting, where models may fail to generalize well to new and unseen instances; (ii) the imbalanced class distribution inherent in many rare disease datasets exacerbates the difficulty, as models may struggle to discern minority classes effectively.

By artificially expanding the training dataset, data augmentation enables CNNs to learn invariant features and nuances, ultimately enhancing the model's ability to generalize and hence improving the model's capability to recognize subtle patterns and features indicative of rare diseases. It is advisable to also conduct comparative analyses between the augmented and non-augmented models to assess the efficacy of data augmentation in improving the robustness and generalization of the CNN model for rare disease image classification.

### 3. Machine Learning-Assisted Statistical Modeling of the Literature (Pertaining to CNN Application for Medical Image Understanding)

Scientific progress relies on the efficient assimilation of published knowledge in order to choose the most promising way forward and to minimize reinvention. But, due to the rapidly evolving nature of the research literature, determining the relevance of an individual report, aggregating and synthesizing multiple reports to derive new insights and finding latent knowledge cannot be efficiently carried out manually. Here, we used machine learning-assisted statistical modeling to search, aggregate, analyze and synthesize the literature on the application of CNN for medical imaging to identify latent and relevant information spread across research articles, conference proceedings and book chapters.

The whole process started by gathering a comprehensive corpus of literature on application of CNN for medical image understanding including journal articles, conference proceedings and book chapters. The gathered large corpus of datasets were preprocessed to handle specialized terminology, abbreviations and language patterns prevalent in the medical imaging literature; to remove noise, ensure consistency and standardize the text, and transform the raw text into a suitable input format. This involved removing irrelevant metadata, handling special characters, standardizing text formatting, tokenizing the text into phrases or words, text normalization, removing stopwords, stemming, lemmatization and spelling corrections while preserving the contextual information and maintain the integrity of the text during preprocessing. Feature mining techniques such as entity recognition, keyword extraction, topic modeling and literature summarization were used to identify trends, patterns and associations and to detect relationships between entities within the existing literature. Language model-based statistical modeling were used to generate coherent summaries, identify method gaps, predict future trends and propose potential solutions based on the patterns and relationships identified during the text mining and analyses stages.

### 3.1. Literature Search Strategy

We used different literature search strategies in which multiple combinations of key words and search engines along with stringent exclusion and inclusion criteria.

#### Frequently used 'key words':

Medical image/imaging, classification, segmentation, convolutional neural network, optimization, architecture, design, hyperparameter tuning, performance metrics, frameworks and data preprocessing.

#### Literature search engines:

We tested 26 different search engines/databases and 5 large language-based AI tools (Supplementary Table S2). Based on coverage-overlap of the tested search engines and specificity metrics, we chose Google Scholar, IEEE Xplore, PubMed and Dimensions as our main search engines to access literatures pertaining to technical resources on the use of CNN for medical imaging.

#### Inclusion and exclusion criteria:

The literature search using Google Scholar and PubMed were largely focused on peer reviewed articles, whereas studies obtained using IEEE Xplore included conference proceeding in addition to peer reviewed articles. All searches were restricted to CNN methods and approaches, particularly focusing on recent developments and improvements that are useful for medical image understanding.

Non-English materials were filtered out as the first exclusion criterion. The exclusion of contents from retracted sources were carried out using RetractionWatch, a database for checking retracted studies and papers. Also, we used Search Smart, a tool that allows researchers to compare the capabilities of most of the conventional search tools, including Google Scholar, IEEE Xplore and PubMed, as an additional exclusion criterion.

### 3.2. Statistical Modeling and Visualization

Machine learning algorithms (implemented as open source, python library or R package) such as non-negative matrix factorization, automated content analysis, Cochrane crowd platform, Rayyan, VOSviewer, Bibliometrix, litsearchr, revtools, wordcloud, wordcloud2, tm and ggplot2 were used for the statistical modeling of the literature and the visualization of the modeling outputs. These tools were used for topic modeling, word frequency counting, network analyses, knowledge graph construction, visualization to uncover latent topics, prevalent themes, method gaps and potential future directions. The statistical modeling involves multiple steps and the functions of each of these packages. For example, we used multiple functions of the "bibliometrix" package, such as "convert2df" to convert the corpus of documents to data frames (as statistical modeling inputs); "biblio-Analysis" for statistical scoring of the data frames; "summary" to see the overall picture of the statistical analyses outputs; "biblioNetwork" to construct networks based on the analyses results; and "plot" for visualization. Similarly, multiple steps were involved with the other packages and tools during the statistical modeling and visualization processes. The detailed steps and scripts for each of these tools and packages can be found at the GitHub repository (the link is provided under the "Data Availability Statement").

## 4. Results from Statistical Modeling

The findings of the statistical modeling, reported in this section, were based on search hits identified using IEEE Xplore, PubMed, Google Scholar and Dimensions. A total of 4609 publications (accessed on 14 June 2023) consisting of 2278 research articles, 938 conference proceedings, 903 book chapters, 470 preprints, 19 edited books and 1 monograph were identified by searching for the keywords "convolutional neural networks" AND "classification" AND ("medical image" OR "medical imaging") in the "Title" AND "abstracts" (which were published from 2006 to 2023). Networks and graphs were visualized using VOSviewer [124,125] and ggplot2 [126]. Bibliometrix (R package) was used to assess publication and citation trends.

IEEE Xplore: The advanced search option of the IEEE Xplore, with the search key words: “classification”, “medical image” and “convolutional neural network” (in the “abstract”) identified 863 unique hits consisting of 687 conference proceedings, 164 journal articles, 7 early access articles and 4 books. Of the 863, 617 citations were published between 2020–2023 (which comprised 484 conference proceedings, 123 journal articles, 7 early access and 3 books) (Figures 4 and S1).

PubMed: Using the advanced search option ((classification [Title/Abstract]) AND (medical image [Title/Abstract])) AND (convolutional neural network [Title/Abstract]), 231 citations were identified. The statistical modeling of the 231 (PubMed) hits were visualized using VOSViewer (Figure 5).

Google Scholar: Using the advanced search option occurring only in the title of the article (screenshot shown next), we identified 212 articles. The statistical modeling of the 212 citations was visualized using VOSViewer (Figure 6).

Transfer learning and segmentation along with the attention mechanisms and incorporation of the transformer seems to be dominant approaches within the combined search hits obtained from IEEE Xplore, PubMed and Google Scholar (Supplementary Figure S2a,b). Medical classification for the diagnostic purpose of COVID-19, brain pathology and breast cancer seem to dominate the literature with regard to application of CNNs for medical image analysis (Supplementary Figures S2 and S3). Widely mentioned metrics to access classification performance include ROC curves and sensitivity and specificity (Supplementary Figure S2). Additionally, the phrases convolutional neural networks, medical image classification, diagnostic imaging, backpropagation, adaptive momentum methods, nonconvex optimization and image interpretation were mentioned more frequently among 141 references cited within the text of this manuscript (Figures 5–7 and Supplementary Figure S4).

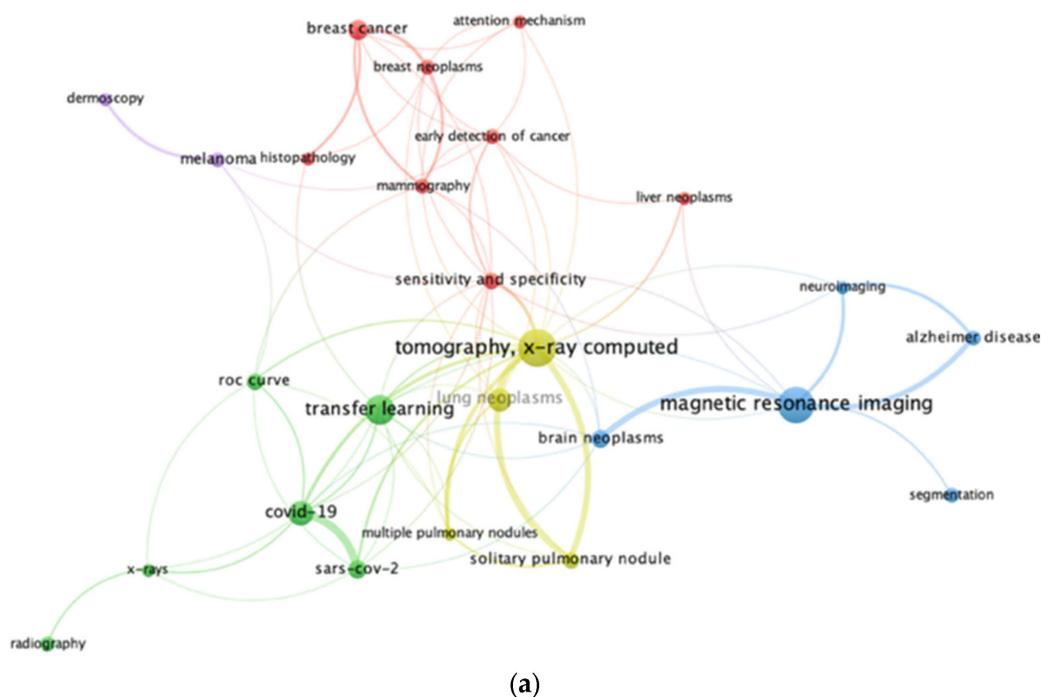
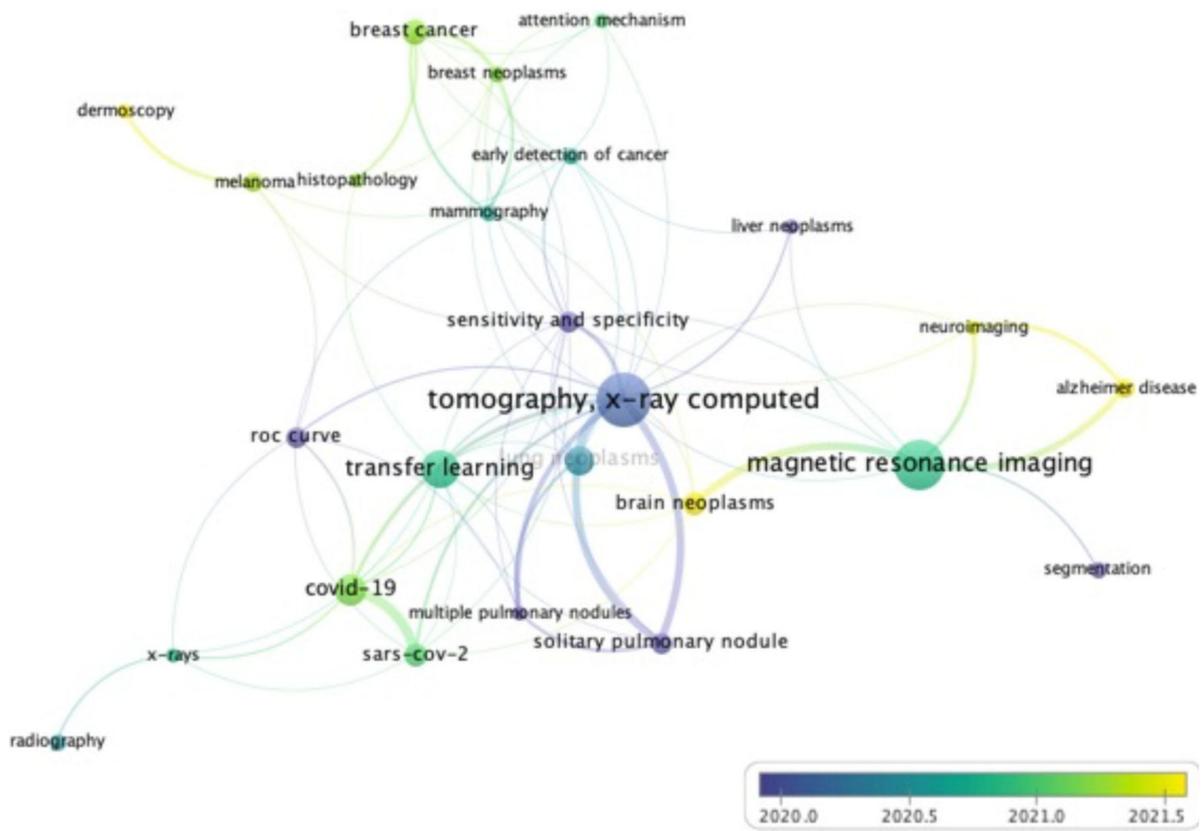


Figure 5. Cont.



(b)

**Figure 5.** Results from statistical modeling of 231 (PubMed) citations. X-ray tomograph (mainly in relation to pulmonary nodules) and magnetic resonance imaging (Alzheimer’s disease and brain neoplasm) seem to be widely analyzed using CNN methods in references indexed by PubMed. Color palettes indicate (a) relationships of studies and (b) publication years. The pandemic does not seem to have significantly changed the patterns of PubMed-indexed publications.

Advanced search 🔍

**Find articles**

with **all of the words**

with the **exact phrase**

with **at least one** of the words

**without** the words

where my words occur

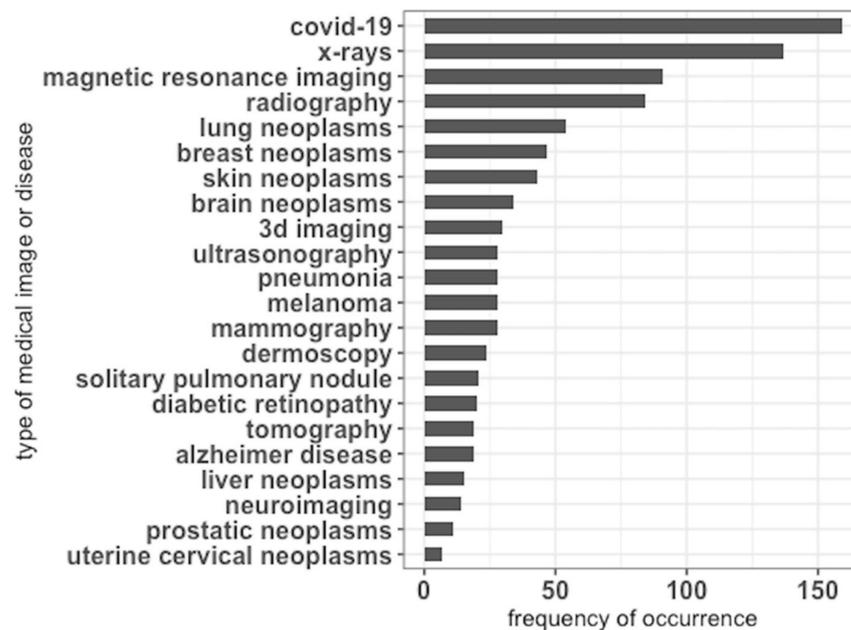
anywhere in the article  
 in the title of the article

**Figure 6.** Cont.



Dimensions: Using the advanced search options and keywords “convolutional neural networks AND classification AND (medical image OR medical imaging)” searched for in “Title AND abstracts”, a total of 4609 hits were found. The search hits include 2278 research articles, 938 proceedings, 903 chapters, 470 preprint, 19 edited books and 1 monograph.

All the search hits obtained using IEEE Xplore, PubMed and Google Scholar were also subsets of the 2278 research articles, 938 proceedings and 903 chapters identified using Dimensions. Key word frequency and word cloud analyses of the 2278 articles 938 proceedings and 903 chapters used to rank popular diagnostic images used as CNN inputs, corresponding diseases, CNN algorithms and evaluation metrics that are applied to understand such medical images (Figures 8 and 9).

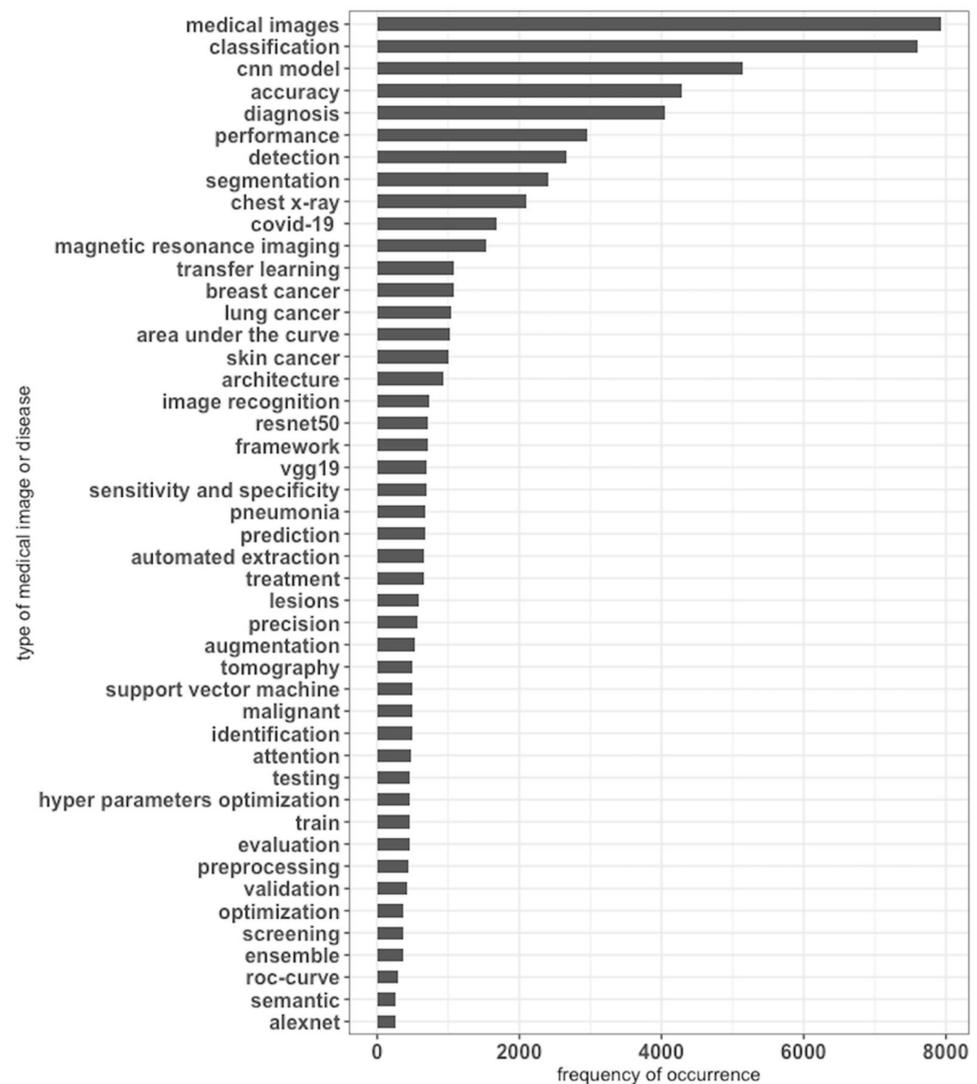


**Figure 8.** Pathological conditions, medical images and performance metrics are more frequently mentioned within the 2278 research articles and the 938 proceeding papers.

The most frequently used medical imaging modalities seem to include X-ray, MRI, radiography, ultrasonography, histopathological staining, CT, tomography and optical endoscopy (Figure 8). These images are used for various medical imaging tasks such as for the detection of COVID-19 [127], lesion detection, image segmentation and image classification in specialties such as radiology, cardiology and gastroenterology (Figure 8).

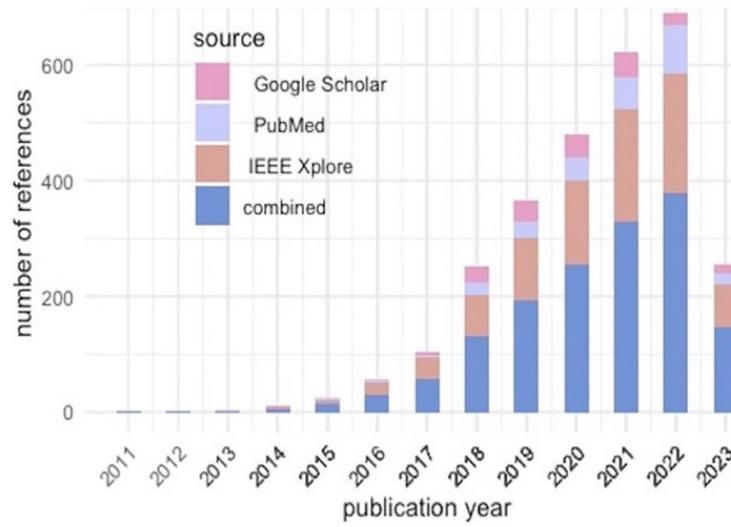
#### Annual distribution of publications identified using search engines

The number of publications per year for the combined search hits showed a steep increase since the start of the COVID-19 pandemic (Figures 10 and S5). This analysis was based on the combined references collected using Google Scholar, PubMed and IEEE Xplore (after excluding duplicates). The number of hits from Google Scholar were small because we focused only on the titles of the articles.

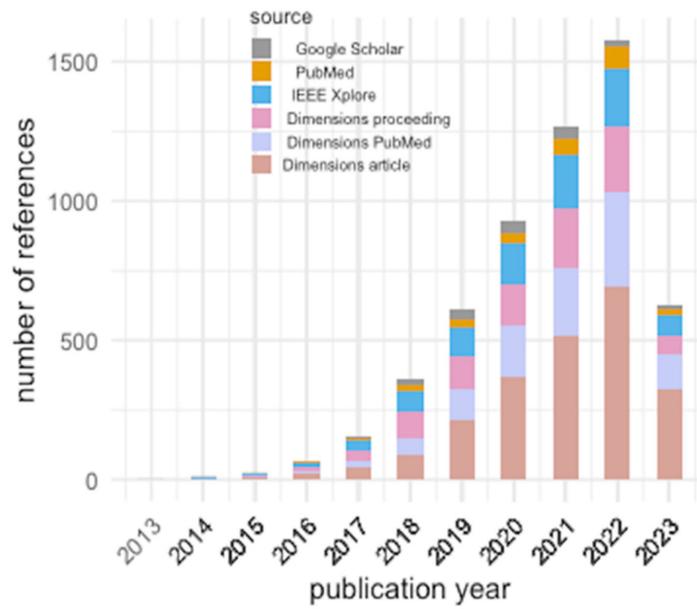


**Figure 9.** Ranking of more frequently used terms and phrases related to methods, diseases, images and metrics that were mentioned in the 2278 research articles, 938 conference proceeding papers and 903 book chapters that are relevant to the application of CNN for medical image understanding.

Bibliometrix v4.1.4 (R package) analyses outputs (Figure 11): Bibliometrix supports bibliographic database files from Dimensions, which also includes all the publication obtained from Google Scholar, IEEE Xplore and PubMed. A steep increase in publications took place after 2018 (probably more propelled by the wide use of CNN for COVID-19 diagnosis). On the other hand, the most cited papers are published around 2016 (Figure 12). An analysis was performed on 2273 research articles published between 2006 and 2023. The summary of the analysis showed that an annual growth rate of 40.6% with an average article age of 2.01 years, 33.57 average citations per article and 7.056 average citations per year per article (Figure 11a). Publications in the journals “Multimedia Tools and Applications”, “IEEE Access”, “Diagnostics” and “Applied Sciences” are highly cited (Figure 13).



(a)



(b)

**Figure 10.** Distribution of publications per year for (a) found using IEEE Xplore, PubMed, Google Scholar and unique combinations of the search hits from the three search engines or databases; (b) identified using Dimensions (which also includes the search hits obtained from IEEE Xplore, PubMed and Google Scholar).

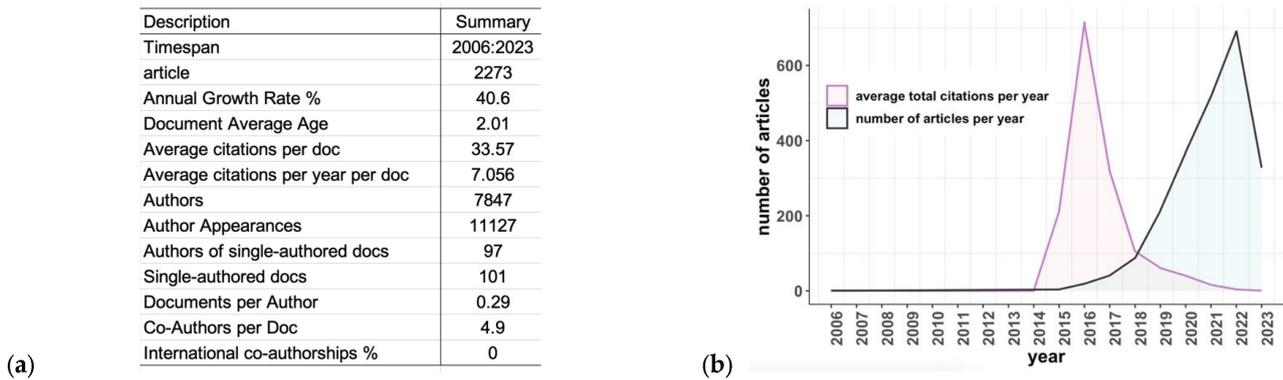


Figure 11. Bibliometrix analyses results showing (a) summary information about 2273 peer reviewed articles on the application of CNN for medical imaging and (b) the annual scientific production and average total citations per year.

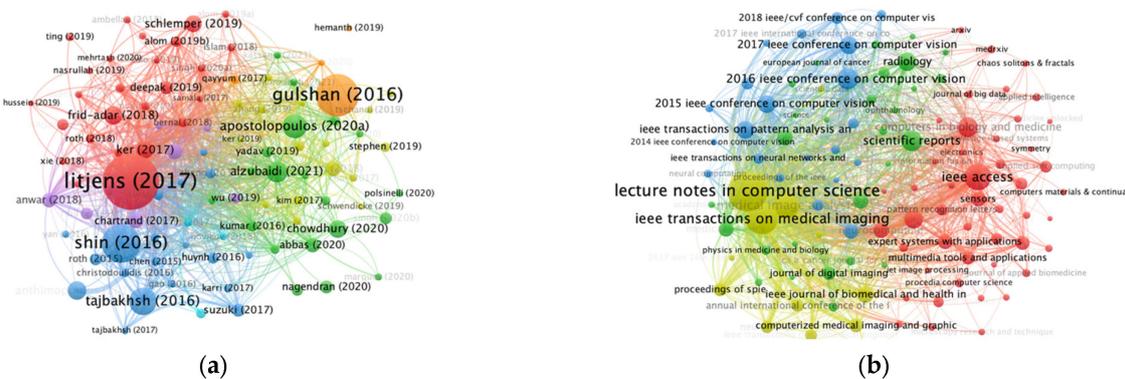


Figure 12. (a) Most of the CNN-related foundational works seem to have been published between 2015 and 2018, which helped for an explosive growth of the field during the pandemic (or it may be that prior works on applying CNN for medical image classifications were not very numerous and that most of the papers published from 2015 to 2019 were disproportionately cited by the flood of papers during the pandemic and in the ensuing years). (b) Co-citation sources, i.e., the number of papers published by the different journals. Node sizes are proportional to the number of papers or proceedings published in that particular journal, and colors indicate either years of publications or inter-citation clustering.

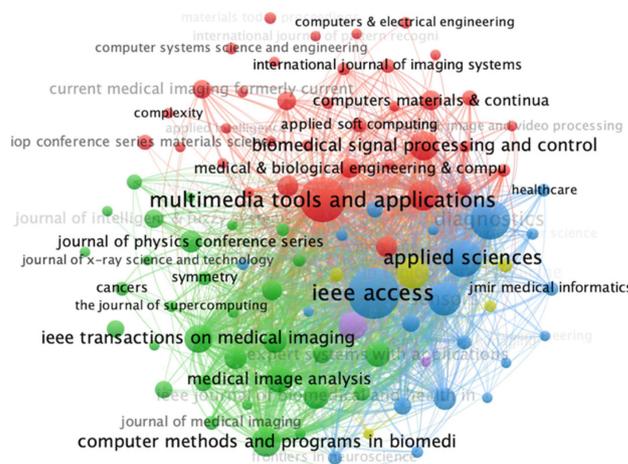
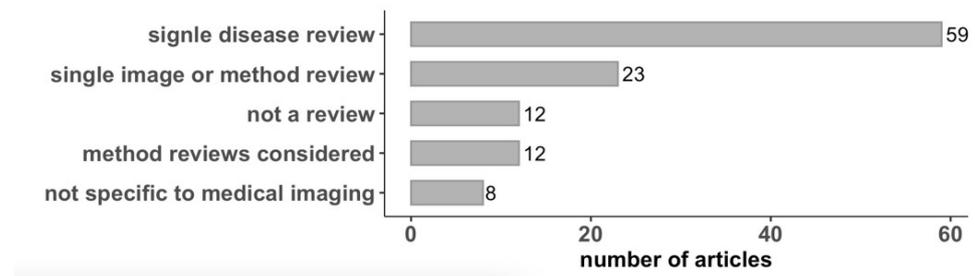


Figure 13. Network showing bibliographic coupling of sources (cross-journal citations). Node size indicates the number of citations, and colors indicated inter-journal citations.

Using Dimensions as a search tool, with the search criteria “Review AND CNN AND MEDICAL AND (IMAGE OR IMAGING) AND CLASSIFICATION” in the titles and abstracts (accessed on 30 July 2023), 115 articles were identified. One non-English article was excluded. Of the 114 articles, 59 were reviews on results of studies focusing on a specific disease, 23 were on single image or method, 12 were not review papers and 8 review papers were not specific to medical imaging (Figure 14). Of the 12 method review papers considered (Figure 14), 8 were broad and shallow introductory or background method reviews and the remaining 4 method review papers were comparable to this study (Table 6).



**Figure 14.** Distribution of the 114 articles identified using Dimensions and using the key words “Review AND CNN AND MEDICAL AND (IMAGE OR IMAGING) AND CLASSIFICATION” in the titles and abstracts.

**Table 6.** Comparison of the current study with other method review papers on medical image understanding. x denotes yes, and xx denotes the comprehensive coverage of alternative and/or improved CNN components or algorithms.

Review Paper	Relevant Datasets	Data Pre-processing	Architecture	Activation Functions	Frame Works	Optimization Methods	Ensemble Techniques	Performance Metrics	Statistical Modeling
Munir et al. [128]		x	x	x	x	x		xx	
Salhi et al. [3]		xx	xx	x	x	x		xx	
Abdou [129]		x	xx	x	x	x		x	
Yao et al. [130]		x	xx	x	x	x		x	
This review	x	xx	xx	xx	xx	xx	x	xx	xx

### Summary of statistical modeling results

The majority of articles were published during the COVID-19 pandemic and afterwards (Figures 8–11). It seems that more imaging data became available during COVID-19, making transfer learning and GAN-based data augmentation less important compared to papers published pre-COVID-19 (Figure 5). The analysis of chest X-rays, histopathology and endoscopic images, diabetic retinopathy and neuroimaging, brain tumor/neoplasm classification and detections utilizing CNNs was identified during and post-COVID-19 (Figures 5–7). Likewise, the use of transformers, image fusion schemes, genetic algorithms and momentum trended more during and after the COVID-19 pandemic (Figures 5–7). Overall, the majority of the medical image classifications were applied for the recognition of pathological conditions and the detection and diagnosis of diseases such as COVID-19 (including pneumonia detection) and lung and breast cancers, as predicted through chest X-ray and CT images.

## 5. Discussion

### 5.1. Highlights of Current Practices

The use of CNNs in medical image understanding has shown significant improvements. CNN models for medical image classifications can be trained from scratch, using

off-the-shelf pretrained models (transfer learning) and/or conducting unsupervised CNN pretraining with supervised fine-tuning [131,132]. Ensembles of different CNN models and combinations of CNN algorithms with transformers, including global spatial attention mechanisms [133], are being explored for the classification of multiple pathological images such as X-ray, MRI, CT and histopathological stains.

### 5.2. Implications for Clinical Practice

Medical image understanding using CNN has shown promising results in various medical domains, including disease classification, tumor segmentation, lesion detection, identifying anatomical location [50,134–136] and diagnosing COVID-19 and metastatic cancer with high classification accuracy [137]. Overall, the use of CNNs in medical image classification has significant implications for clinical practice, including improving the accuracy and speed of diagnosis, and for treatment planning. Ongoing research in these areas aims to advance the field and improve the effectiveness, interpretability and applicability of CNN models in clinical settings.

### 5.3. Gaps in the Current State of CNN Application for Medical Image Understanding

CNNs are being applied to a wide range of medical images [138] including X-ray, MRI, CT, optical coherence tomography and histopathology stains. However, there are still several open questions and challenges regarding the current state of CNN application for medical image understanding. Such gaps include: (i) The problem of interpretability (considering CNN models as black boxes), making it challenging to interpret their decisions and understand the reasoning behind their predictions, which can be a significant barrier to the adoption of CNNs in medical imaging. (ii) The limited availability of annotated medical imaging datasets—the performance of CNNs is highly dependent on the quality and quantity of the training data. (iii) Robustness to the diverse data and pathological variations—medical images can exhibit significant variations due to factors such as different imaging modalities, patient demographics, imaging protocols and disease presentations. Ensuring CNN models' robustness and generalizability across diverse data distributions and pathological variations is an open question. (iv) Finding efficient methods for domain adaptation and transfer learning to enhance the generalizability of CNN models across different medical imaging domains is still challenging. (v) Addressing class imbalance and rare diseases—medical image datasets often suffer from class imbalance, where certain diseases or conditions are underrepresented. Developing techniques to handle class imbalance and effectively train CNN models on rare diseases is an ongoing challenge. (vi) Uncertainty estimation—determining how to effectively incorporate uncertainty estimation into CNN models and providing an error margin for predictions is an ongoing research challenge. (vii) Integration with clinical workflow—the successful integration of CNN models into clinical practice requires addressing workflow-related challenges, including seamless integration with existing medical systems, interpretability in a clinical context and adapting CNN models to real-time decision support systems. (viii) Data privacy and security—medical images contain sensitive patient information, and preserving patient privacy is important. Exploring techniques like federated learning, differential privacy or secure computation to enable CNN training on distributed medical image data while preserving privacy is an open question.

### 5.4. Trends and Future Directions

CNNs are considered a significant technological breakthrough in the field of medical image understanding and are increasingly gaining attention [139,140]. It is evident that CNNs have been successfully employed in medical image recognition, segmentation and classification. The use of class decomposition and transfer learning, synthetic data augmentation, Bayesian and adaptive hyperparameter optimizations and specialized architectures can be used to improve the robustness and performance of CNN models for the classi-

fication of medical images [3,131,132,141–143], including the diagnosis and prognosis of diseases such as COVID-19 [140].

While CNNs have shown promise in medical image understanding, further research is needed to optimize their performance for efficient medical diagnosis and treatment follow-up or evaluation. Some of the future trends and potential improvements of CNN approaches in the field of medical image classification include:

- The development of specialized and efficient CNN architectures (including methods for automatically designing CNN architectures), such as evolving arbitrary CNNs with the goal of better discovering and understanding new effective architectures for robust learning outcomes that are tailored towards learning specific representations.
- Design methods that can be readily used for automatic optimization of CNN architectures for personalized medicine.
- Designing new domain agnostic CNN algorithms which can be used for transfer learning or that can be used for reliable learning from small datasets or learn online, e.g., by combining with reinforcement learning.
- Exploring new activation functions for efficient and robust learning, including on small datasets (to mitigate the issue of labeled data scarcity).
- Aiming to improve feature extraction efficiency of CNNs using multilinear filters [144] and designing accelerators for CNN inferences [145] to improve the speed and accuracy of medical image analysis tasks.
- Designing more efficient 3D CNNs.
- The dynamic selection of misclassified negative samples during training to improve performance and to speed up learning.
- The privacy, data security and prevention of adversarial data poisoning.

These recommendations can help set up more robust, improved, secure and reliable CNN experiments for medical image understanding.

### 5.5. Limitations of the Review

This review focuses on the methodological aspect of CNN as applied for medical image understanding, and hence, it does not include a summary of the findings and results of the literature that are not directly related to developments or advancements of CNN methods.

Additionally, the current literature on the application of CNN for medical image classification has some limitations in scope. First, most research efforts focus on adapting existing CNN architectures with or without transfer learning, rather than designing and optimizing CNNs architectures and approaches specific to medical image classification. Second, there is a lack of research on the amount of data needed to train deep CNN models to achieve high accuracy. Third, while CNNs have shown competitive performance in medical image analysis tasks, such as disease classification, segmentation and detection, there is a lack of post hoc explainability for CNN-based models [146] as well as a need for more research on the use of CNNs for low resource medical image analysis [143]. Even though there are advancements with regard to light-weight CNN architectures for economic GPU-based systems [67,68,147], still more concerted efforts are needed to make automated medical image analysis accessible in real time.

## 6. Conclusions

The focus of this review is on the methodological aspects of CNNs' applications for medical image understanding. It is organized to serve as a resource for practitioners by compiling and presenting improved architectures, popular frameworks, activation functions, model ensemble techniques, hyperparameter optimizations, performance metrics, medical theme relevant datasets and input data preprocessing methods that are important for better CNN designs and to learn robust models.

We also used machine learning (ML) algorithms for the statistical modeling of the literature to uncover latent topics, patterns and prevalent themes, method gaps and potential future directions that were not obvious from individual studies. Our ML-assisted analyses

showed that the COVID-19 pandemic probably stimulated the wide use of CNN for clinical image classifications and disease diagnosis. The COVID-19 problem probably drove the flow of CNN practitioners to the discipline of medical imaging, apparently creating an atmosphere of collaboration with people in the biomedical field. This may be the reason for the drastic increase in journal articles and conference proceedings pertaining to the application of CNN for medical image recognition, analysis, classification and disease detection and diagnosis.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/make6010033/s1>. Figure S1: Results from statistical modeling of 863 citations identified using IEEE Xplore. Network visualized using VOSviewer; Figure S2: Analysis outputs of the combined references (search hits) from IEEE Xplore, PubMed and Google Scholar. Color palettes represent (a) relationships of the corresponding studies and (b) publications years; Figure S3: Systematic analysis of the references cited in this review (and creating a network using VOSviewer with settings: Create a map based on bibliographic data and Co-occurrence counting method for Keywords)); Figure S4: Most frequently mentioned methods among cited references: convolutional neural networks, medical image classification, diagnostic imaging, backpropagation, adaptive momentum methods, nonconvex optimization and image interpretation; Figure S5: Distribution of references per year for (a) the three search engines (IEEE Xplore, PubMed, Google Scholar) and unique combinations of the search hits from three search engines. (b) Obtained from Dimensions and the search hits obtained from IEEE Xplore, PubMed and Google Scholar; Table S1: List of hyperparameters; Table S2: List of search engines. Table S3: Representative classical CNN methods, and their applications on well-known public datasets including metrics used to assess their performance. Table S4: Studies comparing the performances of different CNN methods including metrics used for comparison. Table S5: The efficiency comparison among SOTA CNN methods in various medical image classification tasks (comparison of some widely used CNN architectures in medical image classification).

**Author Contributions:** Conceptualization, F.A.M., K.K.T. and S.M.; methodology, F.A.M. and S.M.; software, F.A.M.; formal analysis, F.A.M.; investigation, F.A.M.; resources, F.A.M., K.K.T., M.J. and S.M.; data curation, F.A.M.; writing—original draft preparation, F.A.M.; writing—review and editing, S.M., K.K.T., B.G.A. and M.J.; visualization, F.A.M. and S.M.; supervision, K.K.T., M.J. and S.M.; project administration, K.K.T., M.J. and S.M.; funding acquisition, F.A.M. and M.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Statistical modeling and visualization scripts are available at: <https://github.com/Foziyaam/Statistical-modeling> (accessed on 2 February 2024).

**Acknowledgments:** F.A.M. was financially and logistically supported by Wolkite University, Wolkite, Ethiopia, and Addis Ababa Science and Technology University, Addis Ababa, Ethiopia.

**Conflicts of Interest:** The authors declare no conflicts of interest. The views, opinions, and findings contained in this report are those of the authors and should not be construed as official Department of the Army positions, policies, or decisions, unless so designated by other official documentation. Citations of commercial organizations or trade names in this report do not constitute an official Department of the Army endorsement or approval of the products or services of these organizations.

## Abbreviations

CNN: convolutional neural network; ML: machine learning; GAN: generative adversarial network; MRI: magnetic resonance imaging; CT: computerized tomography; RNN: recurrent neural network; TP: true positive; TN: true negative; FP: false positive; FN: false negative; MSE: mean squared error; ROC: receiver operating characteristic; AUC: area under the curve; LLM: large language model; AI: artificial intelligence.

## References

- Razzak, M.I.; Naz, S.; Zaib, A. Deep learning for medical image processing: Overview, challenges and the future. In *Classification in BioApps: Automation of Decision Making*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 323–350.
- Salih, O.; Duffy, K.J. Optimization Convolutional Neural Network for Automatic Skin Lesion Diagnosis Using a Genetic Algorithm. *Appl. Sci.* **2023**, *13*, 3248. [[CrossRef](#)]
- Salehi, A.W.; Khan, S.; Gupta, G.; Alabdullah, B.I.; Almjally, A.; Alsolai, H.; Siddiqui, T.; Mellit, A. A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope. *Sustainability* **2023**, *15*, 5930. [[CrossRef](#)]
- Sarvamangala, D.; Kulkarni, R.V. Convolutional neural networks in medical image understanding: A survey. *Evol. Intell.* **2022**, *15*, 1–22. [[CrossRef](#)]
- Cheng, Y.; Zhao, C.; Neupane, P.; Benjamin, B.; Wang, J.; Zhang, T. Applicability and Trend of the Artificial Intelligence (AI) on Bioenergy Research between 1991–2021: A Bibliometric Analysis. *Energies* **2023**, *16*, 1235. [[CrossRef](#)]
- Al Fryan, L.H.; Shomo, M.I.; Alazzam, M.B. Application of Deep Learning System Technology in Identification of Women’s Breast Cancer. *Medicina* **2023**, *59*, 487. [[CrossRef](#)]
- Alaba, S. Image Classification using Different Machine Learning Techniques. *TechRxiv* **2023**. [[CrossRef](#)]
- Chan, H.-P.; Samala, R.K.; Hadjiiski, L.M.; Zhou, C. Deep learning in medical image analysis. *Deep Learn. Med. Image Anal. Chall. Appl.* **2020**, *1213*, 3–21.
- Inamullah; Hassan, S.; Alrajeh, N.A.; Mohammed, E.A.; Khan, S. Data Diversity in Convolutional Neural Network Based Ensemble Model for Diabetic Retinopathy. *Biomimetics* **2023**, *8*, 187. [[CrossRef](#)] [[PubMed](#)]
- Fu, Y.; Lei, Y.; Wang, T.; Curran, W.J.; Liu, T.; Yang, X. Deep learning in medical image registration: A review. *Phys. Med. Biol.* **2020**, *65*, 20TR01. [[CrossRef](#)]
- El-Ghany, S.A.; Azad, M.; Elmogy, M. Robustness Fine-Tuning Deep Learning Model for Cancers Diagnosis Based on Histopathology Image Analysis. *Diagnostics* **2023**, *13*, 699. [[CrossRef](#)]
- Equbal, A.; Masood, S.; Equbal, I.; Ahmad, S.; Khan, N.Z.; Khan, Z.A. Artificial intelligence against COVID-19 Pandemic: A Comprehensive Insight. *Curr. Med. Imaging* **2023**, *19*, 1–18.
- Fehling, M.K.; Grosch, F.; Schuster, M.E.; Schick, B.; Lohscheller, J. Fully automatic segmentation of glottis and vocal folds in endoscopic laryngeal high-speed videos using a deep Convolutional LSTM Network. *PLoS ONE* **2020**, *15*, e0227791. [[CrossRef](#)] [[PubMed](#)]
- Krupička, R.; Mareček, S.; Malá, C.; Lang, M.; Klempíř, O.; Duspivová, T.; Široká, R.; Jarošíková, T.; Keller, J.; Šonka, K. Automatic substantia nigra segmentation in neuromelanin-sensitive MRI by deep neural network in patients with prodromal and manifest synucleinopathy. *Physiol. Res.* **2019**, *68*, S453–S458. [[CrossRef](#)] [[PubMed](#)]
- Lin, X. Research of Convolutional Neural Network on Image Classification. *Highlights Sci. Eng. Technol.* **2023**, *39*, 855–862. [[CrossRef](#)]
- Khan, A.; Sohail, A.; Zahoor, U.; Qureshi, A. A survey of the recent architectures of deep convolutional neural networks. *arXiv* **2019**, arXiv:1901.06032. [[CrossRef](#)]
- Abdelrazik, M.A.; Zekry, A.; Mohamed, W.A. Efficient Hybrid Algorithm for Human Action Recognition. *J. Image Graph.* **2023**, *11*, 72–81. [[CrossRef](#)]
- Jussupow, E.; Spohrer, K.; Heinzl, A.; Gawlitza, J. Augmenting medical diagnosis decisions? An investigation into physicians’ decision-making process with artificial intelligence. *Inf. Syst. Res.* **2021**, *32*, 713–735. [[CrossRef](#)]
- Liu, J.-W.; Zuo, F.-L.; Guo, Y.-X.; Li, T.-Y.; Chen, J.-M. Research on improved wavelet convolutional wavelet neural networks. *Appl. Intell.* **2021**, *51*, 4106–4126. [[CrossRef](#)]
- Chen, L.; Li, S.; Bai, Q.; Yang, J.; Jiang, S.; Miao, Y. Review of image classification algorithms based on convolutional neural networks. *Remote Sens.* **2021**, *13*, 4712. [[CrossRef](#)]
- Tripathi, K.; Gupta, A.K.; Vyas, R.G. Deep residual learning for image classification using cross validation. *Int. J. Innov. Technol. Explor. Eng.* **2020**, *9*, 1525–1530. [[CrossRef](#)]
- Wang, W.; Yang, X.; Li, X.; Tang, J. Convolutional-capsule network for gastrointestinal endoscopy image classification. *Int. J. Intell. Syst.* **2022**, *37*, 5796–5815. [[CrossRef](#)]
- Lim, M.; Lee, D.; Park, H.; Kang, Y.; Oh, J.; Park, J.-S.; Jang, G.-J.; Kim, J.-H. Convolutional Neural Network based Audio Event Classification. *KSII Trans. Internet Inf. Syst.* **2018**, *12*, 2748–2760.
- Wang, W.; Yang, Y.; Wang, X.; Wang, W.; Li, J. Development of convolutional neural network and its application in image classification: A survey. *Opt. Eng.* **2019**, *58*, 040901. [[CrossRef](#)]
- Kao, C.-C. Optimizing FPGA-Based Convolutional Neural Network Performance. *J. Circuits Syst. Comput.* **2023**, *32*, 2350254. [[CrossRef](#)]
- Jain, A.; Singh, R.; Vatsa, M. On detecting GANs and retouching based synthetic alterations. In Proceedings of the 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS), Redondo Beach, CA, USA, 22–25 October 2018; pp. 1–7.
- Wang, T.; Lan, J.; Han, Z.; Hu, Z.; Huang, Y.; Deng, Y.; Zhang, H.; Wang, J.; Chen, M.; Jiang, H. O-Net: A novel framework with deep fusion of CNN and transformer for simultaneous segmentation and classification. *Front. Neurosci.* **2022**, *16*, 876065. [[CrossRef](#)] [[PubMed](#)]

28. Siddique, N.; Paheding, S.; Elkin, C.P.; Devabhaktuni, V. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access* **2021**, *9*, 82031–82057. [[CrossRef](#)]
29. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, 20 September 2018, Proceedings*; Springer Nature: Cham, Switzerland, 2018; pp. 3–11.
30. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV)*, Stanford, CA, USA, 25–28 October 2016; pp. 565–571.
31. Jha, D.; Riegler, M.A.; Johansen, D.; Halvorsen, P.; Johansen, H.D. Doubleu-net: A deep convolutional neural network for medical image segmentation. In *Proceedings of the 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, Rochester, MN, USA, 28–30 July 2020; pp. 558–564.
32. Shen, J.; Tao, Y.; Guan, H.; Zhen, H.; He, L.; Dong, T.; Wang, S.; Chen, Y.; Chen, Q.; Liu, Z. Clinical Validation and Treatment Plan Evaluation Based on Autodelineation of the Clinical Target Volume for Prostate Cancer Radiotherapy. *Technol. Cancer Res. Treat.* **2023**, *22*, 15330338231164883. [[CrossRef](#)] [[PubMed](#)]
33. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B. Attention u-net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999.
34. Farooq, M.; Hafeez, A. Covid-resnet: A deep learning framework for screening of COVID-19 from radiographs. *arXiv* **2020**, arXiv:2003.14395.
35. Shehab, L.H.; Fahmy, O.M.; Gasser, S.M.; El-Mahallawy, M.S. An efficient brain tumor image segmentation based on deep residual networks (ResNets). *J. King Saud Univ.-Eng. Sci.* **2021**, *33*, 404–412. [[CrossRef](#)]
36. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
37. Iandola, F.; Moskewicz, M.; Karayev, S.; Girshick, R.; Darrell, T.; Keutzer, K. Densenet: Implementing efficient convnet descriptor pyramids. *arXiv* **2014**, arXiv:1404.1869.
38. Jégou, S.; Drozdal, M.; Vazquez, D.; Romero, A.; Bengio, Y. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, USA, 21–26 July 2017; pp. 11–19.
39. Khan, S.; Naseer, M.; Hayat, M.; Zamir, S.W.; Khan, F.S.; Shah, M. Transformers in vision: A survey. *ACM Comput. Surv. (CSUR)* **2022**, *54*, 1–41. [[CrossRef](#)]
40. Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z.; Tang, Y.; Xiao, A.; Xu, C.; Xu, Y. A survey on vision transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 87–110. [[CrossRef](#)] [[PubMed](#)]
41. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, 10–17 October 2021; pp. 10012–10022.
42. Hatamizadeh, A.; Nath, V.; Tang, Y.; Yang, D.; Roth, H.R.; Xu, D. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In *Proceedings of the International MICCAI Brainlesion Workshop, Virtual*, 27 September 2021; pp. 272–284.
43. Hatamizadeh, A.; Tang, Y.; Nath, V.; Yang, D.; Myronenko, A.; Landman, B.; Roth, H.R.; Xu, D. Unetr: Transformers for 3d medical image segmentation. In *Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, USA, 3–8 January 2022; pp. 574–584.
44. Dalmaz, O.; Yurt, M.; Çukur, T. ResViT: Residual vision transformers for multimodal medical image synthesis. *IEEE Trans. Med. Imaging* **2022**, *41*, 2598–2614. [[CrossRef](#)] [[PubMed](#)]
45. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv* **2021**, arXiv:2102.04306.
46. Li, Z.; Li, D.; Xu, C.; Wang, W.; Hong, Q.; Li, Q.; Tian, J. TFCNs: A CNN-Transformer Hybrid Network for Medical Image Segmentation. In *Artificial Neural Networks and Machine Learning—ICANN 2022: 31st International Conference on Artificial Neural Networks, Bristol, UK, September 6–9, 2022, Proceedings*; Part IV; Springer: Cham, Switzerland, 2022; pp. 781–792.
47. Zheng, S.; Lu, J.; Zhao, H.; Zhu, X.; Luo, Z.; Wang, Y.; Fu, Y.; Feng, J.; Xiang, T.; Torr, P.H. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, USA, 20–25 June 2021; pp. 6881–6890.
48. Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; Dai, J. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv* **2020**, arXiv:2010.04159.
49. Valanarasu, J.M.J.; Oza, P.; Hacihaliloglu, I.; Patel, V.M. Medical transformer: Gated axial-attention for medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I*; Springer: Cham, Switzerland, 2021; pp. 36–46.
50. Dai, Y.; Gao, Y.; Liu, F. Transmed: Transformers advance multi-modal medical image classification. *Diagnostics* **2021**, *11*, 1384. [[CrossRef](#)]

51. Wang, Z.; Min, X.; Shi, F.; Jin, R.; Nawrin, S.S.; Yu, I.; Nagatomi, R. SMESwin Unet: Merging CNN and Transformer for Medical Image Segmentation. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2022: 25th International Conference, Singapore, 18–22 September 2022, Proceedings, Part V*; Springer: Cham, Switzerland, 2022; pp. 517–526.
52. Khairandish, M.O.; Sharma, M.; Jain, V.; Chatterjee, J.M.; Jhanjhi, N. A hybrid CNN-SVM threshold segmentation approach for tumor detection and classification of MRI brain images. *IRBM* **2022**, *43*, 290–299. [[CrossRef](#)]
53. Pham, Q.-D.; Nguyen-Truong, H.; Phuong, N.N.; Nguyen, K.N.; Nguyen, C.D.; Bui, T.; Truong, S.Q. Segtransvae: Hybrid cnn-transformer with regularization for medical image segmentation. In Proceedings of the 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), Kolkata, India, 28–31 March 2022; pp. 1–5.
54. Dastider, A.G.; Sadik, F.; Fattah, S.A. An integrated autoencoder-based hybrid CNN-LSTM model for COVID-19 severity prediction from lung ultrasound. *Comput. Biol. Med.* **2021**, *132*, 104296. [[CrossRef](#)]
55. Yu, Z.; Lee, F.; Chen, Q. HCT-net: Hybrid CNN-transformer model based on a neural architecture search network for medical image segmentation. *Appl. Intell.* **2023**, *53*, 19990–20006. [[CrossRef](#)]
56. Sun, Q.; Fang, N.; Liu, Z.; Zhao, L.; Wen, Y.; Lin, H. HybridCTrm: Bridging CNN and transformer for multimodal brain image segmentation. *J. Healthc. Eng.* **2021**, *2021*, 7467261. [[CrossRef](#)] [[PubMed](#)]
57. Sangeetha, S.; Mathivanan, S.K.; Karthikeyan, P.; Rajadurai, H.; Shivahare, B.D.; Mallik, S.; Qin, H. An enhanced multimodal fusion deep learning neural network for lung cancer classification. *Syst. Soft Comput.* **2024**, *6*, 200068.
58. Sharif, M.I.; Li, J.P.; Khan, M.A.; Kadry, S.; Tariq, U. M3BTCNet: Multi model brain tumor classification using metaheuristic deep neural network features optimization. *Neural Comput. Appl.* **2022**, *36*, 95–110. [[CrossRef](#)]
59. Haque, R.; Hassan, M.M.; Bairagi, A.K.; Shariful Islam, S.M. NeuroNet19: An explainable deep neural network model for the classification of brain tumors using magnetic resonance imaging data. *Sci. Rep.* **2024**, *14*, 1524. [[CrossRef](#)]
60. Swain, A.K.; Swetapadma, A.; Rout, J.K.; Balabantaray, B.K. Classification of non-small cell lung cancer types using sparse deep neural network features. *Biomed. Signal Process. Control* **2024**, *87*, 105485. [[CrossRef](#)]
61. Morais, M.; Calisto, F.M.; Santiago, C.; Aleluia, C.; Nascimento, J.C. Classification of breast cancer in Mri with multimodal fusion. In Proceedings of the 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI), Cartagena, Colombia, 18–21 April 2023; pp. 1–4.
62. Kaya, M. Feature fusion-based ensemble CNN learning optimization for automated detection of pediatric pneumonia. *Biomed. Signal Process. Control* **2024**, *87*, 105472. [[CrossRef](#)]
63. Abrantes, J.; Bento e Silva, M.J.N.; Meneses, J.P.; Oliveira, C.; Calisto, F.M.G.F.; Filice, R.W. External validation of a deep learning model for breast density classification. *ECR 2023*. [[CrossRef](#)]
64. Diogo, P.; Morais, M.; Calisto, F.M.; Santiago, C.; Aleluia, C.; Nascimento, J.C. Weakly-Supervised Diagnosis and Detection of Breast Cancer Using Deep Multiple Instance Learning. In Proceedings of the 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI), Cartagena, Colombia, 18–21 April 2023; pp. 1–4.
65. Han, Q.; Qian, X.; Xu, H.; Wu, K.; Meng, L.; Qiu, Z.; Weng, T.; Zhou, B.; Gao, X. DM-CNN: Dynamic Multi-scale Convolutional Neural Network with uncertainty quantification for medical image classification. *Comput. Biol. Med.* **2024**, *168*, 107758. [[CrossRef](#)]
66. He, Y.; Gao, Z.; Li, Y.; Wang, Z. A lightweight multi-modality medical image semantic segmentation network base on the novel UNeXt and Wave-MLP. *Comput. Med. Imaging Graph.* **2024**, *111*, 102311. [[CrossRef](#)]
67. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
68. Qureshi, S.A.; Raza, S.E.A.; Hussain, L.; Malibari, A.A.; Nour, M.K.; Rehman, A.u.; Al-Wesabi, F.N.; Hilal, A.M. Intelligent ultra-light deep learning model for multi-class brain tumor detection. *Appl. Sci.* **2022**, *12*, 3715. [[CrossRef](#)]
69. Xiao, J.; Ye, H.; He, X.; Zhang, H.; Wu, F.; Chua, T.-S. Attentional factorization machines: Learning the weight of feature interactions via attention networks. *arXiv* **2017**, arXiv:1708.04617 2017.
70. Guo, H.; Tang, R.; Ye, Y.; Li, Z.; He, X. DeepFM: A factorization-machine based neural network for CTR prediction. *arXiv* **2017**, arXiv:1703.04247.
71. Wang, R.; Fu, B.; Fu, G.; Wang, M. Deep & cross network for ad click predictions. In Proceedings of the ADKDD'17, Halifax, NS, Canada, 13–17 August 2023; pp. 1–7.
72. Watanabe, S.; Hori, T.; Karita, S.; Hayashi, T.; Nishitoba, J.; Unno, Y.; Soplin, N.E.Y.; Heymann, J.; Wiesner, M.; Chen, N. Espnet: End-to-end speech processing toolkit. *arXiv* **2018**, arXiv:1804.00015.
73. Pratap, V.; Hannun, A.; Xu, Q.; Cai, J.; Kahn, J.; Synnaeve, G.; Liptchinsky, V.; Collobert, R. Wav2letter++: A fast open-source speech recognition system. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 6460–6464.
74. Dai, J.J.; Ding, D.; Shi, D.; Huang, S.; Wang, J.; Qiu, X.; Huang, K.; Song, G.; Wang, Y.; Gong, Q. Bigdl 2.0: Seamless scaling of ai pipelines from laptops to distributed cluster. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 21439–21446.
75. Thongprayoon, C.; Kaewput, W.; Kovvuru, K.; Hansrivijit, P.; Kanduri, S.R.; Bathini, T.; Chewcharat, A.; Leeaphorn, N.; Gonzalez-Suarez, M.L.; Cheungpasitporn, W. Promises of big data and artificial intelligence in nephrology and transplantation. *J. Clin. Med.* **2020**, *9*, 1107. [[CrossRef](#)]

76. Jayasinghe, W.L.P.; Deo, R.C.; Ghahramani, A.; Ghimire, S.; Raj, N. Deep multi-stage reference evapotranspiration forecasting model: Multivariate empirical mode decomposition integrated with the boruta-random forest algorithm. *IEEE Access* **2021**, *9*, 166695–166708. [[CrossRef](#)]
77. Nazari, E.; Biviji, R.; Roshandel, D.; Pour, R.; Shahriari, M.H.; Mehrabian, A.; Tabesh, H. Decision fusion in healthcare and medicine: A narrative review. *Mhealth* **2022**, *8*, 8. [[CrossRef](#)]
78. Santoso, I.B.; Adrianto, Y.; Sensusiati, A.D.; Wulandari, D.P.; Purnama, I.K.E. Ensemble Convolutional Neural Networks with Support Vector Machine for Epilepsy Classification Based on Multi-Sequence of Magnetic Resonance Images. *IEEE Access* **2022**, *10*, 32034–32048. [[CrossRef](#)]
79. Liu, N.; Shen, J.; Xu, M.; Gan, D.; Qi, E.-S.; Gao, B. Improved cost-sensitive support vector machine classifier for breast cancer diagnosis. *Math. Probl. Eng.* **2018**, *2018*, 3875082. [[CrossRef](#)]
80. Qureshi, A.S.; Roos, T. Transfer learning with ensembles of deep neural networks for skin cancer detection in imbalanced data sets. *Neural Process. Lett.* **2022**, *55*, 4461–4479. [[CrossRef](#)]
81. Li, X.; Xiong, H.; Chen, Z.; Huan, J.; Xu, C.-Z.; Dou, D. “In-Network Ensemble”: Deep Ensemble Learning with Diversified Knowledge Distillation. *ACM Trans. Intell. Syst. Technol. (TIST)* **2021**, *12*, 1–19. [[CrossRef](#)]
82. Mukherjee, D.; Dhar, K.; Schwenker, F.; Sarkar, R. Ensemble of deep learning models for sleep apnea detection: An experimental study. *Sensors* **2021**, *21*, 5425. [[CrossRef](#)] [[PubMed](#)]
83. SureshKumar, M.; Perumal, V.; Yuvaraj, G.; Rajasekar, S.J.S. Detection of Pneumonia from Chest X-Ray images using Machine Learning. *Concurr. Eng.-Res. Appl.* **2022**, *30*, 325–334.
84. Cui, W.; Liu, Y.; Li, Y.; Guo, M.; Li, Y.; Li, X.; Wang, T.; Zeng, X.; Ye, C. Semi-supervised brain lesion segmentation with an adapted mean teacher model. In *Information Processing in Medical Imaging: 26th International Conference, IPMI 2019, Hong Kong, China, 2–7 June 2019, Proceedings 26*; Springer: Cham, Switzerland, 2019; pp. 554–565.
85. Toda, R.; Oda, M.; Hayashi, Y.; Otake, Y.; Hashimoto, M. Improved method for COVID-19 classification of complex-architecture CNN from chest CT volumes using orthogonal ensemble networks. In *Proceedings of the SPIE Medical Imaging, San Diego, CA, USA, 19–24 February 2023*; p. 124650D.
86. Chen, Y.-M.; Chen, Y.J.; Ho, W.-H.; Tsai, J.-T. Classifying chest CT images as COVID-19 positive/negative using a convolutional neural network ensemble model and uniform experimental design method. *BMC Bioinform.* **2021**, *22*, 147. [[CrossRef](#)] [[PubMed](#)]
87. Thomas, J.B.; KV, S.; Sulthan, S.M.; Al-Jumaily, A. Deep Feature Meta-Learners Ensemble Models for COVID-19 CT Scan Classification. *Electronics* **2023**, *12*, 684. [[CrossRef](#)]
88. Liu, S.; Xie, Y.; Jirapatnakul, A.; Reeves, A.P. Pulmonary nodule classification in lung cancer screening with three-dimensional convolutional neural networks. *J. Med. Imaging* **2017**, *4*, 041308. [[CrossRef](#)] [[PubMed](#)]
89. Bazgir, O.; Ghosh, S.; Pal, R. Investigation of REFINED CNN ensemble learning for anti-cancer drug sensitivity prediction. *Bioinformatics* **2021**, *37*, i42–i50. [[CrossRef](#)]
90. Patane, A.; Kwiatkowska, M. Calibrating the classifier: Siamese neural network architecture for end-to-end arousal recognition from ECG. In *Machine Learning, Optimization, and Data Science: 4th International Conference, LOD 2018, Volterra, Italy, 13–16 September 2018, Revised Selected Papers 4*; Springer: Cham, Switzerland, 2019; pp. 1–13.
91. Wen, L.; Ye, X.; Gao, L. A new automatic machine learning based hyperparameter optimization for workpiece quality prediction. *Meas. Control* **2020**, *53*, 1088–1098. [[CrossRef](#)]
92. Gu, B.; Liu, G.; Zhang, Y.; Geng, X.; Huang, H. Optimizing large-scale hyperparameters via automated learning algorithm. *arXiv* **2021**, arXiv:2102.09026.
93. Liu, Y.; Li, Q.; Cai, D.; Lu, W. Research on the strategy of locating abnormal data in IOT management platform based on improved modified particle swarm optimization convolutional neural network algorithm. *Authorea Prepr.* **2023**. [[CrossRef](#)]
94. Ait Amou, M.; Xia, K.; Kamhi, S.; Mouhafid, M. A Novel MRI Diagnosis Method for Brain Tumor Classification Based on CNN and Bayesian Optimization. *Healthcare* **2022**, *10*, 494. [[CrossRef](#)]
95. Saeed, T.; Loo, C.K.; Kassim, M.S.S. Ensembles of deep learning framework for stomach abnormalities classification. *CMC Comput. Mater. Contin.* **2022**, *70*, 4357–4372. [[CrossRef](#)]
96. AlBahar, A.; Kim, I.; Yue, X. A robust asymmetric kernel function for Bayesian optimization, with application to image defect detection in manufacturing systems. *IEEE Trans. Autom. Sci. Eng.* **2021**, *19*, 3222–3233. [[CrossRef](#)]
97. Thavasimani, K.; Srinath, N.K. Hyperparameter optimization using custom genetic algorithm for classification of benign and malicious traffic on internet of things-23 dataset. *Int. J. Electr. Comput. Eng.* **2022**, *12*, 4031. [[CrossRef](#)]
98. Ozcan, T.; Basturk, A. Performance improvement of pre-trained convolutional neural networks for action recognition. *Comput. J.* **2021**, *64*, 1715–1730. [[CrossRef](#)]
99. Korade, N.B.; Zuber, M. Stock Price Forecasting using Convolutional Neural Networks and Optimization Techniques. *Int. J. Adv. Comput. Sci. Appl.* **2022**, *13*, 378–385. [[CrossRef](#)]
100. Ghawi, R.; Pfeffer, J. Efficient hyperparameter tuning with grid search for text categorization using KNN approach with BM25 similarity. *Open Comput. Sci.* **2019**, *9*, 160–180. [[CrossRef](#)]
101. Sinha, A.; Khandait, T.; Mohanty, R. A gradient-based bilevel optimization approach for tuning hyperparameters in machine learning. *arXiv* **2020**, arXiv:2007.11022.
102. Florea, A.-C.; Andonie, R. Weighted random search for hyperparameter optimization. *arXiv* **2020**, arXiv:2004.01628. [[CrossRef](#)]

103. Nayak, D.R.; Padhy, N.; Mallick, P.K.; Bagal, D.K.; Kumar, S. Brain tumour classification using noble deep learning approach with parametric optimization through metaheuristics approaches. *Computers* **2022**, *11*, 10. [[CrossRef](#)]
104. Passos, L.A.; Papa, J.P. A metaheuristic-driven approach to fine-tune deep Boltzmann machines. *Appl. Soft Comput.* **2020**, *97*, 105717. [[CrossRef](#)]
105. Ergen, T.; Mirza, A.H.; Kozat, S.S. Energy-Efficient LSTM Networks for Online Learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *31*, 3114–3126. [[CrossRef](#)] [[PubMed](#)]
106. Mujahid, M.; Rustam, F.; Álvarez, R.; Luis Vidal Mazón, J.; Díez, I.d.l.T.; Ashraf, I. Pneumonia Classification from X-ray Images with Inception-V3 and Convolutional Neural Network. *Diagnostics* **2022**, *12*, 1280. [[CrossRef](#)] [[PubMed](#)]
107. Subramanian, B.; Muthusamy, S.; Thangaraj, K.; Panchal, H.; Kasirajan, E.; Marimuthu, A.; Ravi, A. A new method for detection and classification of melanoma skin cancer using deep learning based transfer learning architecture models. *Res. Sq.* **2022**, preprint. [[CrossRef](#)]
108. Gaur, L.; Bhatia, U.; Jhanjhi, N.; Muhammad, G.; Masud, M. Medical image-based detection of COVID-19 using deep convolution neural networks. *Multimed. Syst.* **2021**, *29*, 1729–1738. [[CrossRef](#)] [[PubMed](#)]
109. Suresh, V.; Janik, P.; Rezmer, J.; Leonowicz, Z. Forecasting solar PV output using convolutional neural networks with a sliding window algorithm. *Energies* **2020**, *13*, 723. [[CrossRef](#)]
110. Bhandari, N.; Khare, S.; Walambe, R.; Kotecha, K. Comparison of machine learning and deep learning techniques in promoter prediction across diverse species. *PeerJ Comput. Sci.* **2021**, *7*, e365. [[CrossRef](#)] [[PubMed](#)]
111. Kumar, A.; Kim, J.; Lyndon, D.; Fullham, M.; Feng, D. An ensemble of fine-tuned convolutional neural networks for medical image classification. *IEEE J. Biomed. Health Inform.* **2016**, *21*, 31–40. [[CrossRef](#)] [[PubMed](#)]
112. Cifci, M.A.; Hussain, S.; Canatalay, P.J. Hybrid Deep Learning Approach for Accurate Tumor Detection in Medical Imaging Data. *Diagnostics* **2023**, *13*, 1025. [[CrossRef](#)]
113. Kalantar, R.; Lin, G.; Winfield, J.M.; Messiou, C.; Lalondrelle, S.; Blackledge, M.D.; Koh, D.-M. Automatic segmentation of pelvic cancers using deep learning: State-of-the-art approaches and challenges. *Diagnostics* **2021**, *11*, 1964. [[CrossRef](#)]
114. Li, J.; Han, D.; Wang, X.; Yi, P.; Yan, L.; Li, X. Multi-sensor medical-image fusion technique based on embedding bilateral filter in least squares and salient detection. *Sensors* **2023**, *23*, 3490. [[CrossRef](#)] [[PubMed](#)]
115. Boikos, C.; Imran, M.; De Lusignan, S.; Ortiz, J.R.; Patriarca, P.A.; Mansi, J.A. Integrating Electronic Medical Records and Claims Data for Influenza Vaccine Research. *Vaccines* **2022**, *10*, 727. [[CrossRef](#)] [[PubMed](#)]
116. Chlap, P.; Min, H.; Vandenberg, N.; Dowling, J.; Holloway, L.; Haworth, A. A review of medical image data augmentation techniques for deep learning applications. *J. Med. Imaging Radiat. Oncol.* **2021**, *65*, 545–563. [[CrossRef](#)] [[PubMed](#)]
117. Yoo, J.; Kang, S. Class-Adaptive Data Augmentation for Image Classification. *IEEE Access* **2023**, *11*, 26393–26402. [[CrossRef](#)]
118. Takahashi, R.; Matsubara, T.; Uehara, K. Data augmentation using random image cropping and patching for deep CNNs. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 2917–2931. [[CrossRef](#)]
119. Alkhairi, P.; Windarto, A.P. Classification Analysis of Back propagation-Optimized CNN Performance in Image Processing. *J. Syst. Eng. Inf. Technol. (JOSEIT)* **2023**, *2*, 8–15.
120. Feshawy, S.; Saad, W.; Shokair, M.; Dessouky, M. Proposed Approaches for Brain Tumors Detection Techniques Using Convolutional Neural Networks. *Int. J. Telecommun.* **2022**, *2*, 1–14. [[CrossRef](#)]
121. Alsmirat, M.; Al-Mnayyis, N.; Al-Ayyoub, M.; Asma’A, A.-M. Deep learning-based disk herniation computer aided diagnosis system from mri axial scans. *IEEE Access* **2022**, *10*, 32315–32323. [[CrossRef](#)]
122. Wei, R.; Zhou, F.; Liu, B.; Bai, X.; Fu, D.; Li, Y.; Liang, B.; Wu, Q. Convolutional neural network (CNN) based three dimensional tumor localization using single X-ray projection. *IEEE Access* **2019**, *7*, 37026–37038. [[CrossRef](#)]
123. Gowdra, N.; Sinha, R.; MacDonell, S. Examining and mitigating kernel saturation in convolutional neural networks using negative images. In Proceedings of the IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society, Singapore, 18–21 October 2020; pp. 465–470.
124. Van Eck, N.; Waltman, L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* **2010**, *84*, 523–538. [[CrossRef](#)]
125. Yu, Y.; Li, Y.; Zhang, Z.; Gu, Z.; Zhong, H.; Zha, Q.; Yang, L.; Zhu, C.; Chen, E. A bibliometric analysis using VOSviewer of publications on COVID-19. *Ann. Transl. Med.* **2020**, *8*, 816. [[CrossRef](#)]
126. Wickham, H. ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* **2011**, *3*, 180–185. [[CrossRef](#)]
127. Islam, M.Z.; Islam, M.M.; Asraf, A. A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. *Inform. Med. Unlocked* **2020**, *20*, 100412. [[CrossRef](#)] [[PubMed](#)]
128. Munir, K.; Elahi, H.; Ayub, A.; Frezza, F.; Rizzi, A. Cancer diagnosis using deep learning: A bibliographic review. *Cancers* **2019**, *11*, 1235. [[CrossRef](#)] [[PubMed](#)]
129. Abdou, M.A. Literature review: Efficient deep neural networks techniques for medical image analysis. *Neural Comput. Appl.* **2022**, *34*, 5791–5812. [[CrossRef](#)]
130. Yao, X.; Wang, X.; Wang, S.-H.; Zhang, Y.-D. A comprehensive survey on convolutional neural network in medical image analysis. *Multimed. Tools Appl.* **2020**, *81*, 41361–41405. [[CrossRef](#)]
131. Summers, R. Deep convolutional neural networks for computer-aided detection: Cnn architectures dataset characteristics and transfer learning. *IEEE Trans. Med. Imaging* **2016**, *35*, 1285.

132. Abbas, A.; Abdelsamea, M.M.; Gaber, M.M. Detrac: Transfer learning of class decomposed medical images in convolutional neural networks. *IEEE Access* **2020**, *8*, 74901–74913. [[CrossRef](#)]
133. Xu, L.; Huang, J.; Nitanda, A.; Asaoka, R.; Yamanishi, K. A novel global spatial attention mechanism in convolutional neural network for medical image classification. *arXiv* **2020**, arXiv:2007.15897.
134. Khan, A.H.; Abbas, S.; Khan, M.A.; Farooq, U.; Khan, W.A.; Siddiqui, S.Y.; Ahmad, A. Intelligent model for brain tumor identification using deep learning. *Appl. Comput. Intell. Soft Comput.* **2022**, *2022*, 8104054. [[CrossRef](#)]
135. Mahjoubi, M.A.; Hamida, S.; El Gannour, O.; Cherradi, B.; El Abbassi, A.; Raihani, A. Improved Multiclass Brain Tumor Detection using Convolutional Neural Networks and Magnetic Resonance Imaging. *Int. J. Adv. Comput. Sci. Appl.* **2023**, *14*, 406–414. [[CrossRef](#)]
136. Pham, C.-H.; Tor-Díez, C.; Meunier, H.; Bednarek, N.; Fablet, R.; Passat, N.; Rousseau, F. Multiscale brain MRI super-resolution using deep 3D convolutional networks. *Comput. Med. Imaging Graph.* **2019**, *77*, 101647. [[CrossRef](#)]
137. Papandrianos, N.; Papageorgiou, E.; Anagnostis, A.; Feleki, A. A deep-learning approach for diagnosis of metastatic breast cancer in bones from whole-body scans. *Appl. Sci.* **2020**, *10*, 997. [[CrossRef](#)]
138. Serte, S.; Serener, A.; Al-Turjman, F. Deep learning in medical imaging: A brief review. *Trans. Emerg. Telecommun. Technol.* **2022**, *33*, e4080. [[CrossRef](#)]
139. Ahmed, M.; Du, H.; AlZoubi, A. An ENAS based approach for constructing deep learning models for breast cancer recognition from ultrasound images. *arXiv* **2020**, arXiv:2005.13695.
140. Kugunavar, S.; Prabhakar, C. Convolutional neural networks for the diagnosis and prognosis of the coronavirus disease pandemic. *Vis. Comput. Ind. Biomed. Art* **2021**, *4*, 12. [[CrossRef](#)] [[PubMed](#)]
141. Frid-Adar, M.; Diamant, I.; Klang, E.; Amitai, M.; Goldberger, J.; Greenspan, H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* **2018**, *321*, 321–331. [[CrossRef](#)]
142. Singh, S.P.; Wang, L.; Gupta, S.; Goli, H.; Padmanabhan, P.; Gulyás, B. 3D deep learning on medical images: A review. *Sensors* **2020**, *20*, 5097. [[CrossRef](#)] [[PubMed](#)]
143. Agrawal, T.; Gupta, R.; Narayanan, S. On evaluating CNN representations for low resource medical image classification. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 1363–1367.
144. Tran, D.T.; Iosifidis, A.; Gabbouj, M. Improving efficiency in convolutional neural networks with multilinear filters. *Neural Netw.* **2018**, *105*, 328–339. [[CrossRef](#)]
145. Hegde, K.; Agrawal, R.; Yao, Y.; Fletcher, C.W. Morph: Flexible acceleration for 3d cnn-based video understanding. In Proceedings of the 2018 51st Annual IEEE/ACM International Symposium on Microarchitecture (MICRO), Fukuoka, Japan, 20–24 October 2018; pp. 933–946.
146. Hasenstab, K.A.; Huynh, J.; Masoudi, S.; Cunha, G.M.; Pazzani, M.; Hsiao, A. Feature Interpretation using Generative Adversarial Networks (FIGAN): A Framework for Visualizing a CNN’s Learned Features. *IEEE Access* **2023**, *11*, 5144–5160. [[CrossRef](#)]
147. Fielding, B.; Lawrence, T.; Zhang, L. Evolving and ensembling deep CNN architectures for image classification. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–8.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.