

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/375671489>

A survey on cancer detection via convolutional neural networks: Current challenges and future directions

Article in Neural Networks · January 2024

DOI: 10.1016/j.neunet.2023.11.006

CITATIONS

47

READS

718

5 authors, including:



Pallabi Sharma

National Institute of Technology Meghalaya

18 PUBLICATIONS 246 CITATIONS

[SEE PROFILE](#)



Deepak Ranjan Nayak

Malaviya National Institute of Technology Jaipur

76 PUBLICATIONS 2,633 CITATIONS

[SEE PROFILE](#)



Bunil Kumar Balabantaray

National Institute of Technology Meghalaya

99 PUBLICATIONS 532 CITATIONS

[SEE PROFILE](#)



M. Tanveer

Indian Institute of Technology Indore

271 PUBLICATIONS 8,782 CITATIONS

[SEE PROFILE](#)

A Survey on Cancer Detection via Convolutional Neural Networks: Current Challenges and Future Directions

Pallabi Sharma^a, Deepak Ranjan Nayak^b, Bunil Kumar Balabantaray^c, M. Tanveer^{d,*}, Rajashree Nayak^e

^a*School of Computer Science, UPES, Dehradun, 248007, Uttarakhand, India*

^b*Department of Computer Science and Engineering, Malaviya National Institute of Technology, Jaipur, 302017, Rajasthan, India*

^c*Computer Science and Engineering, National Institute of Technology Meghalaya, Shillong, 793003, Meghalaya, India*

^d*Department of Mathematics, Indian Institute of Technology Indore, Simrol, 453552, Indore, India*

^e*School of Applied Sciences, Birla Global University, Bhubaneswar, 751029, Odisha, India*

Abstract

Cancer is a condition in which abnormal cells uncontrollably split and damage the body tissues. Hence, detecting cancer at an early stage is highly essential. Currently, medical images play an indispensable role in detecting various cancers; however, manual interpretation of these images by radiologists is observer-dependent, time-consuming, and tedious. An automatic decision-making process is thus an essential need for cancer detection and diagnosis. This paper presents a comprehensive survey on automated cancer detection in various human body organs, namely, the breast, lung, liver, prostate, brain, skin, and colon, using convolutional neural networks (CNN) and medical imaging techniques. It also includes a brief discussion about deep learning based on state-of-the-art cancer detection methods, their outcomes, and the possible medical imaging data used. Eventually, the description of the dataset used for cancer detection, the limitations of the existing solutions, future trends, and challenges in this domain are discussed. The utmost goal of this paper is to provide a piece of comprehensive and insightful information to researchers who have a keen interest in developing CNN-based models for cancer detection.

Keywords: Automated cancer detection, Medical imaging, Deep learning, CNN, Classification, Segmentation

*Corresponding author

Email addresses: pallabi.sharma@ddn.upes.ac.in (Pallabi Sharma), drnayak.cse@mnit.ac.in (Deepak Ranjan Nayak), bunil@nitm.ac.in (Bunil Kumar Balabantaray), mtanveer@iiti.ac.in (M. Tanveer), rajashree@jisiasr.org (Rajashree Nayak)

1. Introduction

Analysis of abnormal cell growth inside the human body that causes cancer is crucial since it is considered as one of the highly life threatening diseases in human. According to the World Health Organization (WHO), about 10 million cancer-related deaths and 19.3 million new cancer cases were diagnosed in 2020 [1, 2]. It has been estimated that the cancer burden will rise to 28.4 million by 2040 [1, 3]. Table 1 reports the most common cancer types and the death rates per year [1, 2, 3, 4]. The breast cancer and lung cancer have the highest fatality rate compared to other cancer types. Therefore, there is an urgent need to improve the detection procedure to minimise cancer instances and death caused by cancer.

Table 1: Statistics of common cancer types.

Cancer Type	No. of Cases (in million per year)	No. of death Cases (in million per year)
Breast	2.26	0.685
Lung	2.21	1.80
Colorectal	1.93	0.916
Prostate	1.41	-
Skin cancer	1.20	-
Liver	0.90	0.830
Brain	0.33	0.251

The early diagnosis of an affected cell may help to decrease the rate of mortality. To diagnose the same, medical imaging modalities such as magnetic resonance imaging (MRI), computerized tomography (CT), positron emission tomography (PET), X-ray, etc., have been extensively used by medical experts. Each of these modalities provide different information about the affected region and hence, it is imperative to decide on the right type of modalities that can facilitate accurate lesion identification. As human healthcare is one of the highly sensitive area of attention, these diagnosis decision are mostly taken by manual observations by the medical experts. However, manual interpretation of these images to find the presence of abnormalities is tedious, time-consuming, and observer-dependent. Therefore, the design of computer-aided diagnosis (CAD) systems for automated and timely detection of disease has gained remarkable attention in the past few years and is still an active area of research. The CAD system can assist clinicians in the fast and accurate detection of disease [5, 6, 7].

In general, a CAD system takes input as medical images and involves image processing and machine learning (ML) techniques to accurately predict the disease [8]. Such systems were introduced in the early 90's and are still very popular for analyzing medical images [9]. Extraction of relevant features so-called handcrafted features and classification are the two most vital stages in CAD systems that uses traditional ML techniques. However, the selection of suitable features and classification techniques has still remained a major concern in these systems. In contrast, deep learning (DL) based methods learn high-level feature representations automatically from images and

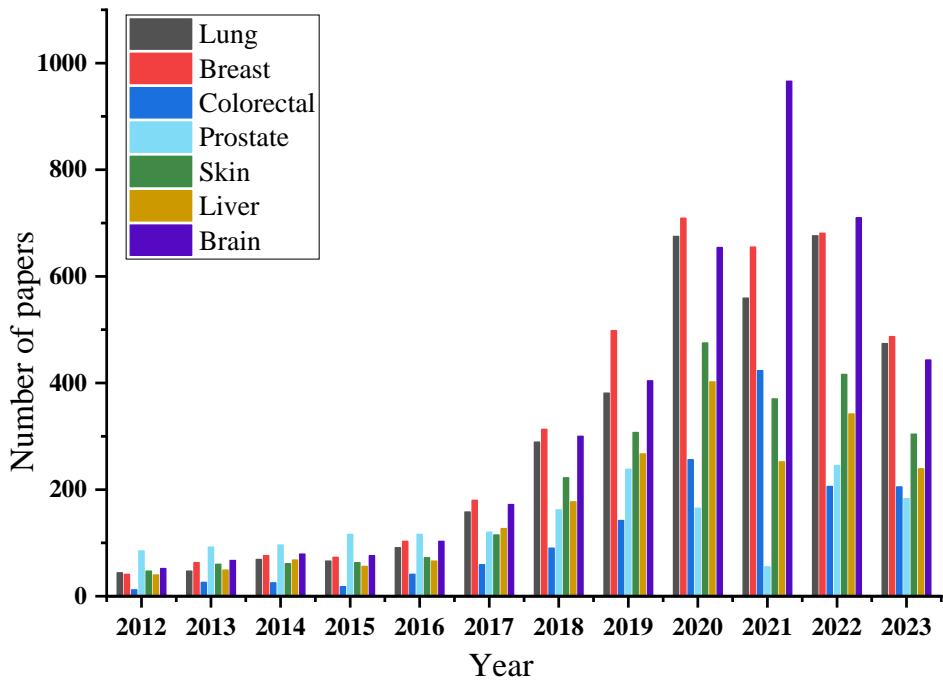


Figure 1: Number of papers published in between 2012-2023 for different cancer detection

have shown significant performance in a wide variety of computer vision applications, including medical image analysis. The most frequently used DL methods for medical image analysis to date are convolutional neural networks (CNNs), which often demand a huge amount of data (e.g., ImageNet) for effective model training [10]. However, the availability of large quantities of medical images is a big concern [11]. Also, the lack of a standard dataset poses problems in the design of DL-based CAD systems for accurate cancer detection. Further, comparing those methods’ performance or reproducing the previous findings is a difficult task due to privacy concerns [12]. Therefore, this has still remained an ongoing research topic.

A scientometric analysis of past publications has been carried out for the duration from January 2012 to May 2023 and it has been observed that since the last decade, researchers in this domain are attempting to solve the problems that arise in expert decision-making. But, a solution close to meet the clinical standard is still in high demand. Figure 1 shows number of published papers from year 2012-2023 for cancer detection in various organs. We can see the sudden growth of paper publication from the year 2017 to date. The growth rate of ML and DL-based techniques in automatic cancer detection in the last decade is illustrated in Figure 2. It can be observed that the DL methods are exploded in popularity in the last couple of years. Among all DL architectures, CNN has been widely adopted for cancer diagnosis through medical images and the number of papers published by CNN and other DL methods is shown in Figure 3. It can be noticed that the number of papers based on only CNNs has

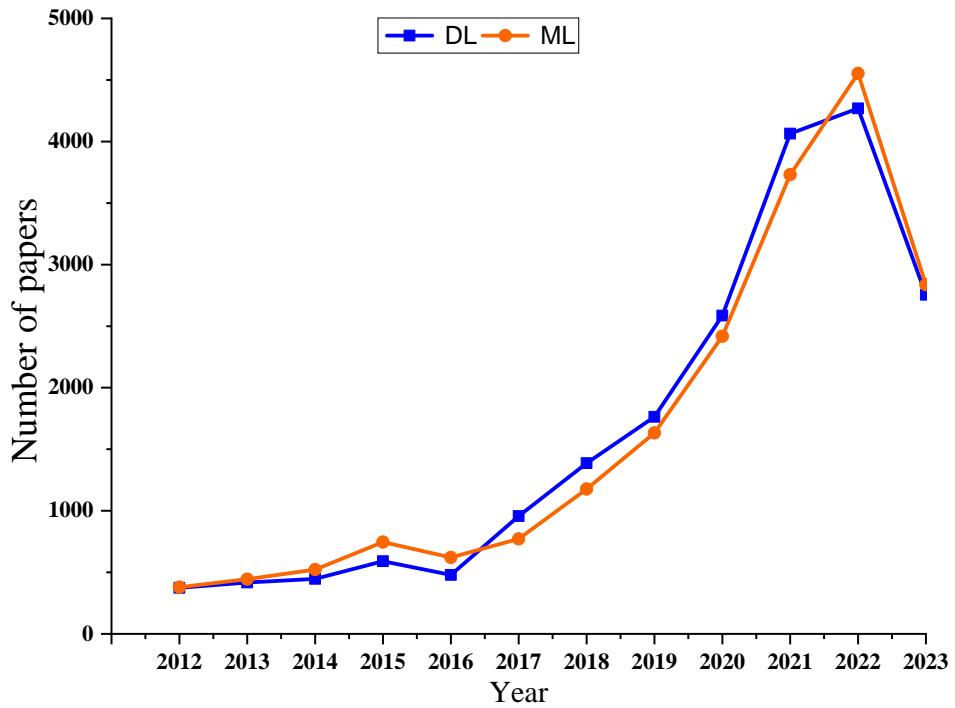


Figure 2: Number of papers published based on ML and DL techniques

been growing exponentially over the last decade, and it is expected that CNN is likely to surpass other DL architectures soon in terms of publication. These pieces of information were collected for a period between 2012 to 2023. The search strategy adopted to select the suitable articles for this survey is discussed in the following section.

1.1. Search strategy

Based on the objectives of this survey, we have reviewed the articles from several scientific databases such as ScienceDirect, Springer, IEEE Xplore, and ACM Digital Library, by searching the terms “cancer detection using machine learning”, “cancer detection using deep learning”, and “cancer detection using convolutional neural network”. Further we filter the articles based on the keyword: “breast cancer classification and segmentation”, “lung cancer classification and segmentation”, “colorectal cancer classification and segmentation”, “skin cancer classification and segmentation”, “prostate cancer classification and segmentation”, “liver cancer classification and segmentation”, and “brain tumor classification and segmentation”. After finding the large number of research articles in each cancer category, we refine them by selecting the years from 2012-2023.

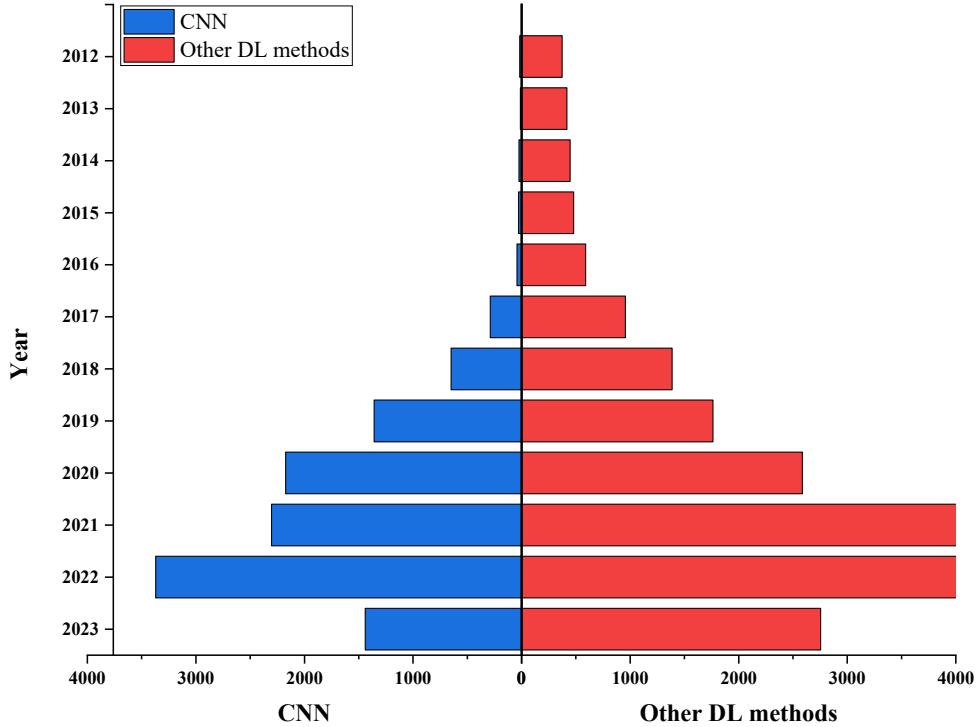


Figure 3: Growth of the papers published on CNN and other DL techniques for the duration 2012-2023

1.2. Contributions

To date, limited review papers have been published on automated cancer detection using DL methods. The majority of the studies have focused on one type of cancer, such as breast cancer [13, 14], brain cancer [15], skin cancer [16], colorectal cancer [17], and very few of them have focused on different variety of cancers [18, 19, 20]. Table 2 lists the summary of published survey papers on cancer detection using DL techniques. While few studies have considered both ML and DL based methods, the information related to dataset are not provided. Further, CNN being the most successful DL method to date has not yet been systematically reviewed in this domain of research. On the other hand, the present study provides a systematic review on the state-of-the-art CNN models for detection of different cancers from medical images. Moreover, a detailed description of the benchmark datasets corresponding to each cancer category is provided. It is worth mentioning here that the number of papers covered in other studies has been counted only by considering the papers listed for each cancer category. Our contributions in this paper are outlined as follows:

- We provide a detailed survey of CNN-based state-of-the-art cancer detection methods published in between January 2012 and May 2023 and highlight their major strengths. In particular, we focus on methods that deal with classification and segmentation tasks.

- We report the benchmark datasets used to validate existing cancer detection methods corresponding to each cancer type, which will help researchers for further development and progress in this direction.
- The survey also discusses the limitation of the existing studies, and current challenges. Further, we aim to provide readers with an overview of future research directions in this domain.

Table 2: Summary of published survey papers on automated cancer detection

Reference	Year	Period	Tasks		Dataset Coverage	Models			No of Cancer Types Covered	No of Papers Covered	Highlights
			C	S		ML	CNN	Other DL			
[20]	2018	2014-2017	✓	✓	Yes		✓	✓	6	76	The article reviewed on breast, lung, skin, prostate, brain and colon cancer, and it provided an overview of different DL architectures.
[18]	2021	2001-2020	✓	✓	No		✓	✓	5	89	This review is basically focused on CNN and the scope to utilize CNN using transfer learning strategy to analyse medical images of brain, heart, breast, lung, and kidney.
[21]	2021	2016-2019	✓	✓	No		✓	✓	3	58	This article provided a review on DL models applicable for medical image classification, segmentation, detection, registration, retrieval, and enhancement.
[19]	2023	2017-2021	✓	✓	No		✓	✓	5	51	Focused on both ML and DL methods used for classification and detection of various cancers. The review is categorized based on cancer type such as brain tumor, cervical, breast, skin and lung.
Ours	2023	2012-2023	✓	✓	Yes		✓		7	98	Provided a comprehensive overview of state-of-the-art CNN architectures and approaches to use them in detecting various cancer types along with the challenges and possible future directions.

C: Classification, S: Segmentation

The remainder of this survey is structured as follows. Section 2 discusses the methods used for automated cancer detection and their achievements. Section 3 provides a brief discussion about CNN. In Section 4, the commonly solved tasks and the performance measures are discussed. Section 5 presents the CNN applications in different types of cancers. Section 6 provides a brief description of open source datasets for each cancer type, and Section 7 discusses the challenges, current trends, and future research directions. Finally, a summary of the survey is presented in Section 8.

2. Automatic cancer detection methods

This section presents the commonly used state-of-the-art methods for automated cancer detection, their pros and cons, and their application in a few important tasks. Figure 4 illustrates the basic steps involved in automated cancer detection from medical images using conventional ML and DL methods.

2.1. Traditional machine learning based automated cancer detection methods

There has been a significant stride in the design of CAD systems using machine learning in the past few years. These systems in general follow a pipeline of multiple

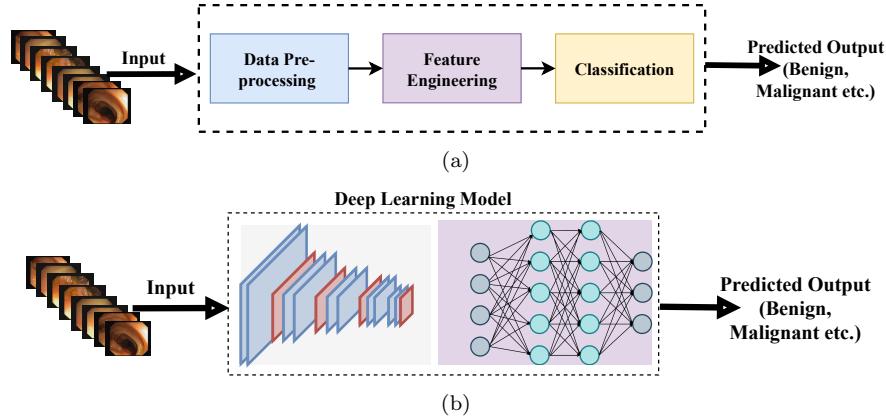


Figure 4: Pipeline of automated cancer detection system using (a) conventional machine learning, and (b) deep learning approach.

stages such as pre-processing, feature extraction, feature selection, and classification. Data pre-processing is an essential stage in many medical image analysis tasks. To perform accurate image analysis, it is indispensable that the images should not contain any unwanted information. Although the technology for medical imaging has significantly improved, the presence of noise and poor contrast can still lead to a false diagnosis. The pre-processing includes various procedures such as gray-scale transformation, noise reduction, restoration, enhancement of the images, removal of background, etc. For instance, in the case of mammography images, pectoral muscles appear in the region of interest. Therefore, a cropping procedure has been adopted to discard the unnecessary region [22]. Yang *et al.* [23] used the N4 bias correction algorithm to remove artifacts and applied thresholding to separate valid imaging regions from the background. Firmino *et al.* [24] used curvature flow filter to remove noise from lung CT images. Perrot *et al.* [25] used a magnification tool using the stochastic gradient descent optimization method to enhance and reveal the subtle motion present in a medical image. Santos and Mascarenhas [26] used pixel-wise non-local means approach to remove Gaussian noise and employed block-matching to remove Rayleigh and fisher-tippett noise. From the literature, it has been observed that a variety of noise removal techniques have been widely used in many medical image analysis to pre-process the images. Since poor image quality is a factor that degrades the performance of any medical image processing task, image enhancement techniques have also been employed frequently to improve the quality of the images, thereby, enhancing the performance. Among all such techniques, histogram equalization has been used mostly to improve the quality of the low-contrast images [27, 28]. Feature extraction is the most crucial stage in machine learning based automated methods. A notable progress has been made in the past years to improve the efficiency of this stage. Features can be of different types such as statistical features like mean, standard deviation, skewness, kurtosis, etc., geometrical features like area perimeter, circularity, equivalent diameter, roundness, etc., texture features, color fea-

tures, multi-resolution features, etc. Although most of the feature descriptors have been developed for computer vision tasks on natural images, such descriptors have been enormously harnessed in medical image analysis problems such as classification and lesion detection [29, 30]. The most common feature extraction techniques include local binary pattern [31], histogram of oriented gradients [32], gray level co-occurrence matrix [22], discrete cosine transform [33], scale-invariant feature transform [34], discrete wavelet transform [35], curvelet transform [36], etc. The features derived using these methods are named hand-crafted or hand-engineered features. However, certain features obtained using different feature extraction methods may be redundant or irrelevant for a particular task, leading to the curse of dimensionality and a decrease in performance. Feature selection techniques are essential to avoid these issues by selecting the most suitable representatives of the data. These techniques facilitate understanding data, lowering computation time and avoiding the issue of the curse of dimensionality. The three methods most frequently employed are filter, wrapper, and embedded methods. In order to select only the high-ranked features, filter methods select the intrinsic properties of features using statistical measures, and the selected features are then used to train the predictor. In contrast, wrappers and embedded methods optimize the objective functions to find a feature subset that gives the highest predictor performance. Embedded methods perform feature selection and training of the algorithm in parallel. In addition, a few dimensionality reduction methods such as principal component analysis, linear discriminant analysis, etc., have been extensively applied in the medical image analysis tasks [37, 38]. The choice of a good feature descriptor for a specific task in medical image analysis heavily relies on domain knowledge and has still remained a challenging task. Hence, feature learning based methods have recently gained considerable attention. In the final stage, a wide variety of classifiers such as K-nearest neighbor, naive Bayes, decision tree, random forest, support vector machine (SVM), feed-forward neural network, extreme learning machine, etc., have been extensively used to classify medical images into different categories. The critical challenge in this stage lies in deciding an appropriate classifier that efficiently utilizes the handcrafted features and results in better performance.

2.2. Deep learning based automated cancer detection methods

DL techniques enable end-to-end solutions and facilitate high-level feature learning from the input data without human intervention, achieving dramatic success in a wide range of computer vision tasks like classification, object detection, and segmentation. Also, it has the ability to handle large-scale data. In recent years, several DL-based methods such as stacked auto-encoders, deep belief networks, restricted Boltzmann machines, recurrent neural networks (RNN), convolutional neural networks (CNN), fully convolutional networks (FCN), have been extensively used for analyzing medical images, including automated cancer detection [39, 20]. Among all these DL architectures, CNNs have been mostly used and found effective for automated cancer detection using medical images.

3. Convolutional neural network

CNN belongs to a family of neural networks and is considered to be the most successful deep neural network model. It learns hierarchical features using multiple layers. Figure 5 shows the basic architecture of CNN which mainly comprises the convolutional layer, ReLU layer, pooling layer, batch normalization, and fully connected (FC) layer.

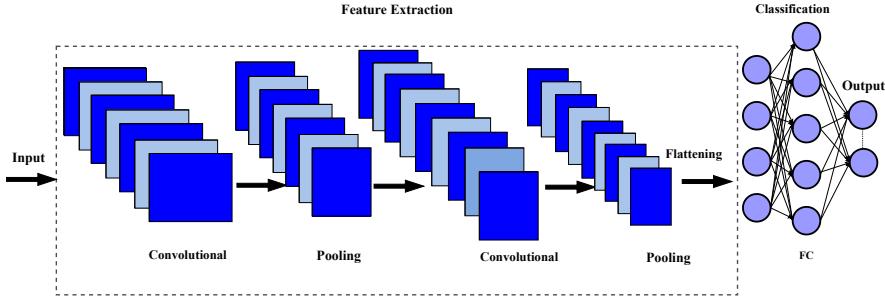


Figure 5: Pipeline of a basic CNN architecture. It usually comprises of convolutional layers, pooling layers followed by fully connected (FC) layers at the end

Convolutional layer: The convolutional layer of CNN plays a vital role in the process of feature extraction. The convolution operation is applied over the input image using a collection of learnable filters, resulting in a diverse set of feature maps. In this layer, individual neurons establish connections with localised regions within the input feature map. For an input image of size $I_h \times I_w \times I_d$, the convolutional layer produces an output image of size $I_h^{new} \times I_w^{new} \times I_d^{new}$ by setting the parameters such as number of filters (n_k), filter size (F), zero padding (P) and stride (S) which can be defined as follows.

$$\left. \begin{aligned} I_w^{new} &= \frac{I_w - F + 2P}{S} + 1, \\ I_h^{new} &= \frac{I_h - F + 2P}{S} + 1, \\ I_d^{new} &= n_k. \end{aligned} \right\} \quad (1)$$

Pooling layer: The pooling layer facilitates the down-sampling of feature maps along spatial dimensions, resulting in a smaller dimensional output [40]. This layer does not demand any parameters to learn. For an input image of size $M \times N$, the output of the pooling layer $O_w \times O_h$ is computed as

$$\left. \begin{aligned} O_w &= \frac{M - w}{S_w} + 1, \\ O_h &= \frac{N - h}{S_h} + 1, \end{aligned} \right\} \quad (2)$$

where O_w and O_h are the width and height of the image, w and h are the width and height of the window, and S_w and S_h are the strides in the direction of width and

height, respectively.

Batch normalization layer: Training of deep neural networks becomes complicated as the distribution of each layer’s input changes during training. Batch normalization standardizes the inputs to a layer using a mini-batch, allowing the model to train faster and decrease variances. It can be applied to either the activation of a prior layer or inputs directly.

Rectified linear unit (ReLU) layer: ReLU is one of the most widely used transfer functions in CNN which introduces non-linearity [41]. It forces all the negative values to zero and preserves the positive values which is defined as

$$\varphi(x) = \begin{cases} x, & x \geq 0, \\ 0, & x < 0. \end{cases} \quad (3)$$

Fully connected (FC) layer: FC layer is usually used at the end in a CNN architecture. In many architectures, multiple FC layers are used. Each neuron in this layer shares its weights with all other neurons of its preceding layer. The requirement of a large number of learnable parameters is one of the significant drawbacks of this layer, which leads to high computational costs during training. Therefore, reducing the number of parameters of FC layers and limiting the use of FC layers are important criteria to develop an efficient CNN model for the classification of medical images.

Like traditional ML-based methods, pre-processing can aid in performance improvements in CNN-based methods. However, the goal of end-to-end learning of CNN from raw data makes pre-processing less popular in CNN-based CAD systems. Additionally, CNN-based models can handle complex and high-dimensional data, including medical images, making them less reliant on traditional feature engineering and pre-processing steps. Further, CNNs have a high computational demand, and the inclusion of pre-processing step might lead to an escalation in computing workload and a deceleration of the training and inference procedures. The pre-processing is still found to be necessary in some cases. However, the choice of pre-processing depends on the specific task and data. The literature on automated cancer detection shows that the most common pre-processing step is data augmentation [42, 43]. In addition, image resize and normalization have been extensively used. Recently, in a few studies, image enhancement techniques have been applied before feeding the images into the CNN model.

3.1. Popular convolutional neural network architectures

Based on these basic blocks, a variety of architectures have been proposed over the years to solve large-scale image classification and segmentation tasks. Figure 6 depicts a few popular CNN architectures and their important components. A brief discussion on such architectures are given below.

3.1.1. VGGNet

The VGGNet architecture is well-known for its consistent structure, and it is distinguished by the application of small convolutional filters of size 3×3 [44]. The

VGG16 and VGG19 are two notable variants of the VGGNet architecture that are distinguished by their number of learnable parameters and layers. For instance, VGG16 consists of sixteen weight layers, of which thirteen are convolutional layers and three are FC layers as shown in Figure 6a. On the other hand, VGG-19 is slightly deeper than VGG-16 and has 16 convolutional layers and three FC layers. The development of deeper CNNs was significantly impacted by the VGGNet, which also laid the groundwork for succeeding CNN designs such as ResNet. Despite its reputation for simplicity, VGGNet necessitates a considerable allocation of computational resources owing to its intricacy and reliance on 3×3 filters. Due to its simple structure, it has been extensively used in the medical image classification task.

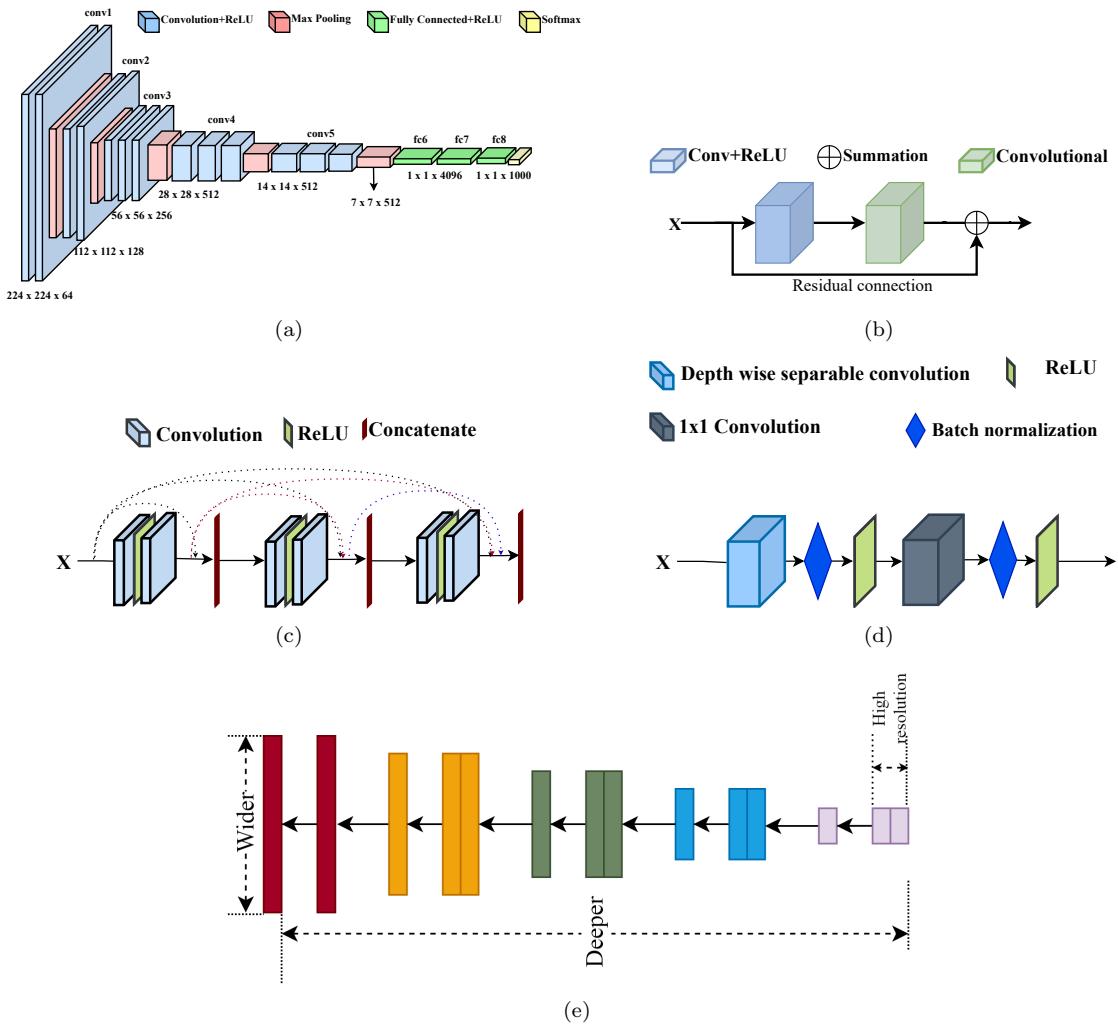


Figure 6: Illustration of CNN architectures and its crucial components: (a) VGGNet architecture [44], (b) residual block, (c) dense Block, (d) convolutional block of MobileNet, and (e) compound scaling of EfficientNet architecture [45].

3.1.2. ResNet

As CNN gets progressively deeper, the information from the input or the gradient computed by the CNN network vanishes between the hidden layers and sometimes washes out by the time it reaches the end or the beginning of the network, causing a new research problem [46]. To solve these problems, He *et al.* [47] proposed an architecture named ResNet and introduced the concept of skip connection. In the conventional feed-forward neural networks, the L^{th} layer transmit the input x through the non-linear function $F(x)$ and produces the output y . Then the output y works as the input to the next layer, resulting the following layer transition as

$$y = F(x). \quad (4)$$

The newly introduced skip-connection bypasses the non-linear transformations with an identity function which can be expressed as

$$y = F(x) + x. \quad (5)$$

The advantage of including this type of skip connection is that any layer that degrades architecture performance can be bypassed. The architecture of convolutional block with skip connection in the of ResNet is shown in Figure 6b. In this architecture, the residual connection does not undergo attenuation or gradient multiplication with activation. Therefore, it is a gradient of unity. Thus, the residual connection makes it feasible to carry forward information to the end of the model and backpropagate the gradient without causing it to disappear.

3.1.3. DenseNet

With an aim of further improvement in the skip connection, Huang *et al.* [48] proposed the DenseNet architecture. The motivation behind introducing this network is to ensure the maximum information flow between the layers in the network. This network architecture connects all layers directly with each other, provided feature-map sizes should be matched and influences the same to all other layers. Each layer acquires some additional inputs from all previous layers and passes on its feature-maps to all successive layers and preserves the feed-forward nature [49, 48]. The underlying concept of the DenseNet architecture is shown in Figure 6c. Unlike ResNet, it does not combine features through addition, instead, it combines features by concatenation before they are passed to the next layer. The DenseNet architecture not only solves the vanishing gradients problem, but it also enables feature reuse, creating the network extremely parameter-efficient.

3.1.4. MobileNet

The application of CNN in real-world scenarios increases the demand for lightweight models. ResNet and DenseNet are efficient in terms of performance but they are power-hungry and computationally expensive. Therefore, they may not be easily deployed in edge devices. Considering this requirement, a lightweight CNN model called

MobileNet was proposed, which uses a depth-wise separable convolutional layer in place of the regular convolutional layer [50]. The convolutional block of MobileNet is shown in Figure 6d. MobileNet-v1 has only 4.2 million learnable parameters, while VGG-16 has 138 million, which is comparably lower than ResNet and DenseNet as well. Less number of learnable parameters makes the MobileNet model fast and energy-efficient for mobile devices. Even though MobileNet performs better than other models in terms of size and speed, the accuracy is still limited. This has led to the development of a new model named EfficientNet.

3.1.5. *EfficientNet*

Increasing the CNN depth or width arbitrarily, or using higher input image resolution for training and evaluation, is the standard technique for model scaling to improve accuracy. Nevertheless, arbitrarily scaling the network dimension increases accuracy, it requires manual fine-tuning with a negligible performance improvement. To solve this issue, Tan and Le [45] introduced the EfficientNet that scales the width, depth, and resolution of the network as shown in Figure 6e with a set of fixed scaling coefficients. The idea behind this compound scaling technique is that a larger input image requires a network with more layers and channels to adequately handle the increased resolution of the larger image. To boost the performance of EfficientNet, inverted bottleneck residual blocks and squeeze-and-excitation blocks were used.

3.1.6. *U-Net*

U-Net is a FCN proposed specifically for image segmentation task [51]. Because of the fully-convolution property, it has a large number of feature maps in the up-sampling path. The architecture is two-way as shown in Figure 7a; the first path is the encoder path which extracts the important features from the images and the other path is a decoder which is a symmetric expanding path that enables localization by upconvolution. There are three components: downsampling, bottlenecking, and upsampling. Each block has a convolutional layer, activation function (with batch normalization), and max pooling. After each pooling layer, U-Net doubles the number of features. This is performed to fulfill the image segmentation task. The upsampling layer of the model composes of a deconvolution layer and activation function. U-Net does not contain any fully connected layer which leads to the model being adaptable to any image size.

3.1.7. *SegNet*

The SegNet architecture is proposed by Badrinarayanan *et al.* [52] for semantic segmentation. The encoder of SegNet is identical to VGG-16 architecture, where the it contains 13 convolutional layers. The key innovation of SegNet is its use of max-pooling indices during the encoding stage to help with the up-sampling process in the corresponding decoder during the decoding stage. These indices are used to store the locations of the max-pooling operation, and are then used to perform up-sampling

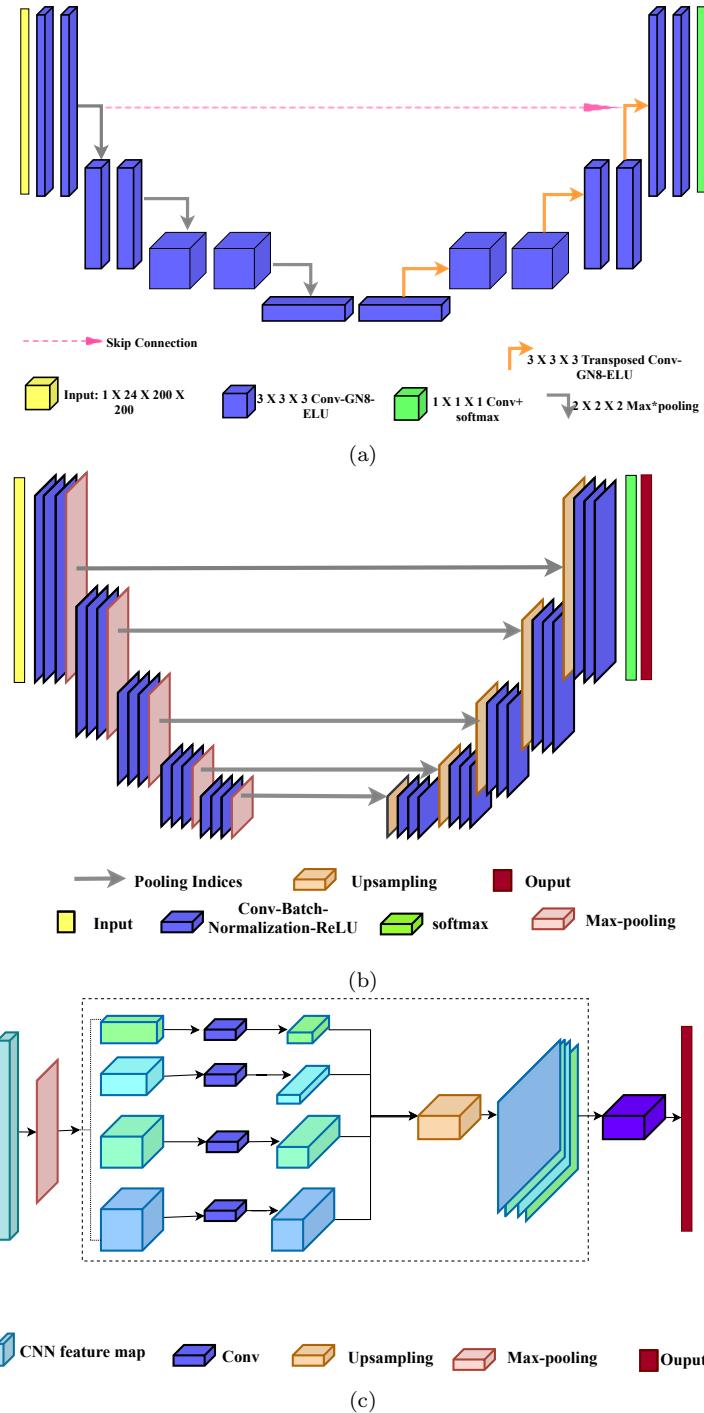


Figure 7: Illustration of popular CNN architectures used for medical image segmentation: (a) U-Net, (b) SegNet, and (c) PSPNet.

during the decoding stage. This helps to preserve spatial information during the up-sampling process and improve the overall performance of the segmentation.

3.1.8. PSPNet

Pyramid Scene Parsing Network (PSPNet) [53] is introduced to capture multi-scale contextual information efficiently, allowing for more accurate pixel-wise segmentation. PSPNet follows the common encoder-decoder architecture. The encoder extracts features from the input image, while the decoder produces the final segmentation mask. The core innovation in PSPNet is the pyramid pooling module (PPM). Instead of using a single fixed-size pooling layer to downsample feature maps, PPM applies pooling operations at multiple spatial scales (pyramid levels). These levels capture context information at different granularities, from fine details to global context. By aggregating features from multiple scales, PSPNet achieves better segmentation performance. Additionally PSPNet employs atrous convolutions (also known as dilated convolutions) in the encoder part to capture multi-scale features with a larger receptive field without increasing the computational cost significantly. While it is powerful in capturing multi-scale context, it can be computationally intensive. Depending on the implementation and the chosen pyramid levels, it may require significant computational resources.

3.1.9. DeepLab

DeepLab is a family of deep learning architectures and are known for their ability to produce high-quality pixel-wise segmentation masks. To capture information at multiple scales, DeepLab employs spatial pyramid pooling. This technique helps the model in understanding both fine-grained and coarse-grained details in the image. The first DeepLab [54] architecture introduces the concept of atrous (dilated) convolutions and atrous spatial pyramid pooling (ASPP). Atrous convolutions allow for capturing multi-scale contextual information, while ASPP captures information at multiple scales. DeepLabv2 [54] is expanded upon DeepLab by incorporating fully connected conditional random fields (CRFs) as a post-processing step. CRFs help in refining segmentation boundaries and improving segmentation accuracy. DeepLabv3 [55] further refines the ASPP module by adding depth-wise separable convolutions, which reduces computation cost while maintaining performance. It also introduces the concept of atrous spatial pyramid pooling with image-level features (ASPP with atrous image pooling) to incorporate global context information. DeepLabv3+ combines the encoder-decoder architecture with atrous convolutions, ASPP, and skip connections. It achieves state-of-the-art performance by capturing both fine and coarse details. The encoder extracts features, and the decoder produces the final segmentation mask. Moreover, DeepLab addresses class imbalance by using techniques like class weighting during training.

3.2. Approaches to use convolutional neural networks architectures

There are various ways to use CNN for the analysis of medical images. First, a CNN model is designed using multiple layers and then trained from scratch using the medical image dataset. Although this approach is very popular, it has a limitation in

that the model may lead to overfitting due to the unavailability of large medical image datasets. Second, the transfer learning concept is adopted in which the pre-trained CNN models (trained on large-scale datasets such as ImageNet) are fine-tuned on the medical image dataset. This approach has been widely applied in medical imaging as it eliminates the need for large datasets. The pre-trained CNN models have also been used as a feature extractor in many studies, where the extracted features are further fed to standard classifiers for classification [43]. This approach does not require the training of the CNN model.

3.3. Attention mechanisms in convolutional neural networks

The human visual system has a remarkable capacity to focus on salient regions in a complex scene. Inspired by this, researchers instigate the use of attention mechanisms in computer vision [56]. These mechanisms have achieved dramatic success in a diverse computer vision tasks such as image classification, object detection, action recognition, face recognition, semantic segmentation, image generation, and medical image analysis. The integration of such mechanisms with CNN architectures has recently become a common practice to improve the performance. This is mainly due to their capability to emphasize on essential features while suppressing the irrelevant ones. Identifying subtle patterns or lesions in various medical images is frequently necessary to discover abnormalities or diseases. The cancerous lesions are present in specific regions of an image and often closely resemble their background. These critical regions of the image can be highlighted by attention mechanisms, helping in the precise localization of disease, tumors, and other anomalies. Additionally, the ability of attention maps to provide insight into a model’s reasoning for a particular decision is yet another essential advantage of attention mechanisms. In medical applications, interpretability is critical because doctors must comprehend the underlying assumptions behind a model’s diagnosis or recommendation. The most commonly used attention mechanisms are spatial attention and channel attention, which capture the feature inter-dependencies along the spatial and channel dimension, respectively. In recent years, a number of channel and spatial attention mechanisms and their combinations, has been proposed, which includes squeeze and excitation (SE) [57], bottleneck attention module (BAM) [58], convolutional block attention module (CBAM) [59], efficient channel attention (ECA) [60], global context (GC) [61], triplet attention [62], joint attention [63], channel split dual attention (CSDA) [64], etc. The effectiveness of such techniques have also been studied over various medical image analysis tasks such as classification, segmentation, lesion detection, etc. Roy *et al.* [65] utilized SE attention module to improve performance of CNN for brain tumor segmentation and whole brain semantic segmentation. Dutta and Nayak [64] integrated CSDA with a backbone network for efffective brain tumor detection. In addition, several recent studies utilized attention mechanisms for detection of various cancers such as colorectal cancer [66, 67, 68], prostate [69] liver [70], lung [71], breast [72], etc. On the other hand, self-attention [73, 74] has caught remarkable

attention from researchers due to its usefulness in many vision tasks. It is considered as a spatial attention mechanism which helps to capture global information. This attention plays a key role in designing vision transformer (ViT) models [75], which have demonstrated remarkable performance in a wide range of computer vision tasks.

4. Problem types and performance Measures

The most commonly solved medical image analysis tasks using CNN include classification, lesion detection, segmentation, and registration. This survey mainly focuses on classification and a few related segmentation tasks.

4.1. Classification

Classification is one of the most crucial problems in the domain of medical imaging which primarily aims to classify medical images into different categories, thereby assisting doctors in verifying their diagnosis. CNN's have been explored the most in the image classification task compared to other DL methods. For instance, Zhang *et al.* [76] proposed a deep CNN model with pre-trained ResNet-50 for skin cancer classification. Yap *et al.* [77] used two variations of CNN (LeNet and FCN-AlexNet) for the classification of breast cancer lesions. Nayak *et al.* [78] designed a lightweight CNN model for multi-class brain abnormalities detection using MR images. Wang *et al.* [79] used pre-trained CNN models via transfer learning concept for multi-classification of endoscopic colonoscopy images.

4.2. Segmentation

Segmentation is also a common task in medical image analysis which is currently being solved using CNN. It is used for identifying anatomical structures such as an organ, vessel, or tumor lesion. Accurate image segmentation is essential as it is easier to analyze and extract significant information from segmented images. Although there exist several segmentation algorithms in the literature [24, 7, 6, 80], their performance is still limited to meet real-time requirements. Deep learning techniques mainly CNN has shown their success in many segmentation tasks and therefore, have been attracting attention to the segmentation of medical images. The existing literature reveals that U-Net is the most dominating CNN model used for segmentation of medical images [81, 77, 82, 83, 84]. Few other CNN models have also been proposed for segmentation purposes. For instance, Rajchl *et al.* [85] used Naive learning based CNN, DeepCut network for brain and lung segmentation. Yang *et al.* [5] used RNN to segment multiple anatomical structures simultaneously, including the fetus, gestational sac, and placenta. Guo *et al.* [86] used a deep CNN-based method to segment images for contouring soft tissue sarcoma lesions using multi-modal images. Hou *et al.* [87] used sparse convolution auto-encoder (SCAE) to detect nuclei in tissue from histopathology images.

4.3. Performance measures

There are various performance metrics used for the evaluation of classification and segmentation methods. This section presents a few commonly adopted measures.

4.3.1. Performance measures for classification task

The assessment of a classification model's performance is based on the ratio of correctly and incorrectly predicted test samples. The confusion matrix reveals the relationship between predicted and actual values. It mainly comprises four elements such as true-positive (TP), false-positive (FP), true-negative (TN), and false-negative (FN). These four elements are used to calculate the following most common performance measures:

- Accuracy: It represents the ratio of the number of correct predictions to all observations.

$$\text{Accuracy} = \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|}. \quad (6)$$

- Precision: It is the ratio of the number of correct positive predictions to total positive predictions.

$$\text{Precision} = \frac{|TP|}{|TP| + |FP|}. \quad (7)$$

- Sensitivity or Recall: It is the ratio of the number of correct positive predictions to actual positive observations.

$$\text{Recall} = \frac{|TP|}{|TP| + |FN|}. \quad (8)$$

- Specificity: It defines the number of correct negative predictions out of all negative observations.

$$\text{Specificity} = \frac{|TN|}{|TN| + |FP|}. \quad (9)$$

- F1-score: The F1-score is the harmonic mean of both precision and recall.

$$\text{F1-Score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}. \quad (10)$$

- Classification-error-rate: It defines the number of negative predictions out of all observations.

$$\text{Classification-error-rate} = \frac{|FP| + |FN|}{|TP| + |TN| + |FP| + |FN|}. \quad (11)$$

4.3.2. Performance measures for segmentation task

The measures used to evaluate segmentation techniques are primarily divided into two groups: supervised and unsupervised. In supervised evaluation criteria, the efficiency of a particular segmentation technique is evaluated by comparing a reference image with the output image generated by the segmentation models. Unsupervised performance measures are used when reference images are not available. In this situation, inter-segment color contrast, intra-segment similarity, shape, and entropy are used as parameters to evaluate the segmentation techniques.

- Dice-coefficient or F1-score: It is the measure of overlap between ground truth and output segmented image.

$$\text{Dice-coefficient} = \frac{2 \times |TP|}{2 \times |TP| + |FN| + |FP|}. \quad (12)$$

- Jaccard-coefficient: It represents the area ratio between the region that is shared by both the ground truth and the predicted segmentation mask, relative to the union of predicted and ground truth mask.

$$\text{Jaccard-coefficient} = \frac{|TP|}{|TP| + |FN| + |FP|}. \quad (13)$$

- Hausdorff-distance: Hausdorff-distance is the distance measured from the edge of the ground truth image to the edge of the segmented region.

$$\text{Hausdorff-distance} = \text{Max}(H_{SG}, H_{GS}), \quad (14)$$

where G is the ground truth image generated by experts and S is the image generated by the segmentation model.

5. Applications of convolutional neural networks in various cancer detection

This section presents an overview of recent studies on cancer detection using CNN. The survey has been organized based on the application areas.

5.1. Breast cancer

Breast cancer is the most common cancer among all cancer types. Several imaging techniques have been successfully used to detect breast cancer which include mammography [12, 88, 89], ultrasound [90, 91], MRI [92, 93], and histopathology [94, 80, 95, 96, 97] images. Plenty of papers have been published on automated breast cancer diagnosis using CNN models. However, there exist no large publicly available datasets to date. From the paper we surveyed, it has been seen that researchers are mostly using popular pre-trained networks such as ResNet, GoogleNet, VGG16,

VGG19 [95, 89, 97, 12] and fine-tuned them on breast cancer images due to the unavailability of large-sized datasets. Also, it can be noticed that many of them used private datasets to validate their proposed methods. In [88, 90], authors developed customized CNNs intending to improve the performance. Wu *et al.* [88] customized the ResNet architecture and considered high-resolution input images for classification task. For segmentation task, Li *et al.* [94] and Wang *et al.* [90] developed customized 3D-CNN similar to U-Net architecture. Recently, Li *et al.* [93] proposed a dual-path CNN architecture for simultaneous segmentation and detection of breast cancer, while for classification of breast tumor in ultrasound image Luo *et al.* [91] employed channel attention to combine feature maps extracted from U-Net segmentation results with the features from original images. In a recent study, He *et al.* [98] introduced a hybrid CNN-transformer network (HCTNet) for breast lesion segmentation. Table 3 includes the summary of the recent work in breast cancer detection that applies CNN as a detection model. Despite the improved performance obtained by several architectures for automated breast cancer analysis, further study is required to enhance the system's performance for real-time clinical applications.

5.2. Lung cancer

Despite improvements in techniques to detect lung cancer in the early stage, it has been found to be the second most common cancer after breast cancer. A pulmonary nodule pertains to malformation in the lung tissue and is reported predominantly in people with lung cancer. Lung cancer is broadly categorized into two categories: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). Several CNN-based lung cancer detection methods have been published in recent years. Table 4 describes the recent CNN-based automated methods for lung cancer detection. The table primarily specifies the following details of a paper: type of problem solved, imaging modalities used, CNN architectures adopted, datasets used and the performance achieved. It has been observed that a variety of imaging methods such as CT [102, 107, 71, 114], PET [105], X-ray [103, 107], and whole slide imaging (WSI) [106] have been used to detect pulmonary nodules. But, among all, CT has been most widely used to analyze lung nodule [115, 114]. Further, it has been noticed that majority of the papers adopted pre-trained CNN architectures such as VGG19 [107] and ResNet [109] or proposed customized CNN models [71] to accomplish classification task. While for segmentation task, Seg3DNet [103], U-Net [81], and mask R-CNN [110] have been widely used and obtained satisfactory results. Another work on lung tumor segmentation embedded a hybrid attention mechanism module with DenseNet architecture to enrich the multiscale features [112]. However, there still exists abundant scope for potential improvements.

5.3. Colorectal cancer

The term colorectal cancer is a combination of colon cancer and rectal cancer. Colon cancer begins in the large intestines (colon), and rectal cancer begins in the

Table 3: Summary of existing CNN models used for breast cancer detection. S → Segmentation, C → Classification.

Authors	Year	Application	Task S C	Modality	CNN Architecture	Dataset	Performance
Bayramoglu <i>et al.</i> [99]	2016	Classification of breast cancer	✗ ✓	Histopathology	Customized CNN	BreaKHis: 7,909 images	Accuracy= 0.8463
Xi <i>et al.</i> [12]	2018	Classification and localization of calcification's and masses	✗ ✓	Mammogram	DCNN, ResNet, GoogLeNet and VGGNet	Merged dataset: (MIAS: 322 images and DDSM: 1644 images)	Accuracy= 0.9242 (by GoogLeNet)
Li <i>et al.</i> [94]	2019	Mitosis segmentation	✓ ✗	Histopathology	SegMitosis	MITOSIS: 648 images, TUPAC16 dataset: 115 images	Precision=0.846, F1-score= 0.8021 (on MITOSIS) and Precision=0.6685, F1-score= 67.28 (on TUPAC16)
Qi <i>et al.</i> [100]	2019	Detection of malignant tumors and solid nodules in breast mass	✗ ✓	Ultrasoundography images	Mt-Net, Sn-Net	Private dataset: 8145 images	Accuracy= 0.9013, Sensitivity= 0.9354, Specificity= 0.8318.
Hu <i>et al.</i> [92]	2020	Benign and malignant breast cancer lesions detection	✗ ✓	Multiparametric MRI	VGG19+SVM with fusion	Private data: 927 images	AUC= 0.87 on Feature-Fusion.
Wang <i>et al.</i> [90]	2020	Automated cancer detection with 3D reconstruction	✓ ✗	ABUS (Automated Breast Ultrasound)	Customized 3D U-	Private dataset: 1415 volumes	Sensitivity= 0.95, Precision= 0.50, F1-score= 0.61, Dice= 0.58.
Wu <i>et al.</i> [88]	2020	Breast cancer classification for high resolution input	✗ ✓	Mammography	Customized ResNet	Private dataset: 1,001,093 images	AUC= 0.876.
Yadavendra and Chand [95]	2021	Breast cancer classification	✗ ✓	Histopathology	Xception	Private dataset: 274,524	Precision= 0.90, Recall= 0.90, F1-score= 0.90.
Yu <i>et al.</i> [101]	2021	Breast abnormalities classification	✗ ✓	Mammography images	ResNet-50	Merged dataset: MINI-MIAS:322 images and INbreast: 330 images	Accuracy= 0.9574
Saber <i>et al.</i> [89]	2021	Classification of mammogram images as benign, malignant and normal classes	✗ ✓	Mammogram images	Inception V3, ResNet50, VGG-19,VGG-16, and Inception-V2ResNet	Private dataset: 322 images	Accuracy= 0.9896, Sensitivity= 0.9783, Specificity= 0.9913, F1-score= 0.9766, and AUC= 0.995 (VGG16)
Liu <i>et al.</i> [97]	2022	Benign and malignant classification	✗ ✓	Histopathology images	AlexNet	BreaKHis: 7909 images, IDC: 162 images and UCSB: 58 images	Accuracy = 0.9841 (on BreaKHis), 0.8631 (on IDC), 0.961 (on UCSB).
Li <i>et al.</i> [93]	2022	Breast mass segmentation	✓ ✓	Mammography	DualCoreNet	CBIS-DDSM: 2478 images, INBreast: 410 images	Dice= 93.69 (on IN-Breast) and 92.17 (on CBIS-DDSM)
Luo <i>et al.</i> [91]	2022	Breast tumor segmentation-to-classification	✓ ✓	Ultrasound images	Channel attention with ResNet and U-Net	Private dataset: segmentation dataset: 292 image, classification dataset: 1702 images	Accuracy= 0.9078, sensitivity= 0.9118 and AUC= 0.9549
He <i>et al.</i> [98]	2023	Cancerous lesion segmentation	✓ ✗	Ultrasound images	Transformer based U-Net	Three Private dataset: total 1263 images	Dice= 0.82 Accuracy= 0.9694, Jaccard= 0.7184, Recall= 0.8214 and Precision= 0.8324.

rectum [141, 142]. Colorectal cancer usually starts as polyps, which may be benign (non-cancerous). Polyps are small clumps of cells and may become cancerous over time. Hence, it is imperative to identify this cancer at an early stage. Stool tests, colonoscopy, sigmoidoscopy, wireless capsule endoscopy, CT colonoscopy, and biopsy are the available tests for colorectal cancer detection. Among all, colonoscopy has been used mostly for screening and biopsy has been used as a confirmatory test for detection. CNN has become a method of choice for the detection of colorectal cancer in recent years. Table 5 describes the recent CNN based techniques adopted for colorectal cancer detection. From the fourteen reviewed papers, it is observed that six

Table 4: Summary of CNN models used for lung cancer detection. S → Segmentation, C → Classification.

Authors	Year	Application	Task		Modality	CNN Architecture	Dataset	Performance
			S	C				
Dou <i>et al.</i> [102]	2016	Pulmonary nodule detection	✗	✓	CT	Fusion of 2D-CNN and 3D-CNN	LIDC: 669 images	Sensitivity= 0.677
Gerard <i>et al.</i> [103]	2018	Pulmonary fissure detection	✓	✓	X-ray, CT	FissureNet and Seg3DNet	Private dataset: 5307 images	Precision= 0.98, AUC= 0.98
Zhang <i>et al.</i> [81]	2018	Lung cysts segmentation	✓	✗	CT	U-Net using recursive learning	Private dataset: 183 volumes	Dice= 0.8265
Jiang <i>et al.</i> [104]	2019	Lung tumor segmentation	✓	✗	CT	Incremental Multi-model Recurrent Neural Network with Residual Network	LIDC: 2669 images, TCIA: 377 images	Accuracy= 0.89 (on TCIA), Accuracy= 0.72 (on LIDC)
Moitra and Mandal [105]	2020	Automatic staging and grading of NSCLC	✓	✓	PET, CT	CNN	NSCLC: 52,073 images	Accuracy= 0.96
Wang <i>et al.</i> [106]	2020	Classification of non-small cell and small cell lung cancer	✗	✓	WSI	CNN+FCN with feature aggregator	TCGA: 500 images, SUCC: 939 images,	Accuracy= 0.82 (on TCGA) and 0.973 (on SUCC)
Ibrahim <i>et al.</i> [107]	2021	Lung disease detection	✗	✓	X-Ray, CT	VGG19, ResNet151V2	Generated dataset by [108]: 20,000 images	Accuracy = 0.9805, F1-score= 0.9824, AUC= 0.9966.
Tong <i>et al.</i> [109]	2021	Automatic lung nodule binary classification	✗	✓	CT	3D-ResNet	LIDC-IDRI: 1,018 images Private dataset: 14415 images	Accuracy = 0.9129 (on LIDC-IDRI), Accuracy = 0.8470 (on private dataset)
Xu <i>et al.</i> [110]	2021	Pulmonary segmentation	✓	✗	CT	Mask R-CNN	Private dataset: 1265 images, LIDC: 36 images	Accuracy= 0.97, Dice= 0.98 (on private dataset), Accuracy= 0.98, Dice= 0.99 (on LIDC)
Causey <i>et al.</i> [111]	2022	Lung cancer predictions	✗	✓	CT	Ensemble of 3D-CNN and Spatial pyramid pooling	National Lung Screening Trial (NLST) cohort: 1449	AUC= 0.892
Hu <i>et al.</i> [112]	2021	Segmentation of lung tumors	✓	✗	CT	DenseNet with hybrid channel and spatial attention	RIDER dataset: 15,419 images	Accuracy= 0.9461
Huang <i>et al.</i> [113]	2022	Benign-Malignant lung nodule classification	✗	✓	CT	Self-supervised transfer learning based on domain adaptation (SSTL-DA) 3D-CNN	LIDC-IDRI dataset: 1,018 clinical thoracic CT	Accuracy= 0.9107 and AUC = 0.9584.
Xu et al. [114]	2022	Multi-category non-small cell lung cancer classification	✗	✓	CT	CNN with attention	Two private datasets	Accuracy= 0.9524 (dataset 1) and 0.9814 (dataset 2)
Xiao et al [71]	2023	Binary classification of lung adenocarcinoma (ADC) and lung squamous cell carcinoma (LSCC)	✗	✓	CT	Customized CNN	TCIA dataset	Accuracy= 0.99 (ADC classification) and 0.9167 (LSCC classification)

papers were focused on solving only classification tasks, and six papers focused only on polyp segmentation tasks while the rest two papers explored both classification and segmentation. VGG19, ResNet50, ResNet101, and GoogleNet are the most dominant CNN architectures for polyps classification [79, 118, 116, 119, 121]. But, due to similar structures of different types of polyps, these models still lead to misclassification. Further, these models require data augmentation for effective fine-tuning because of the unavailability of large-scale data which leads to high training overhead. For segmentation of polyps region, U-Net is the most popular model [43, 122, 123, 125]. But, there is no standard model yet proposed. From studies [43, 122, 123, 68], it has been observed that U-Net mispredicts the boundary pixels of polyps and the performance varies based on the size of the polyps. Considering these drawbacks of

Table 5: Summary of the existing CNN models for colorectal cancer detection. S → Segmentation, C → Classification.

Authors	Year	Application	Task		Modality	CNN Architecture	Dataset	Performance		
			S	C						
Eycke <i>et al.</i> [116]	2018	Segmentation of glandular epithelium	✓	✗	Microscopic images	VGGNet with residual connection	GlaS dataset: 165 images	F1-score= 0.893, Dice= 0.908		
Akbari <i>et al.</i> [117]	2018	Detection of frames that contain polyps	✗	✓	Colonoscopy	CNN with binarized weight	Asu Mayo test clinic dataset: 17,574 images	Accuracy= 0.9028		
Kather <i>et al.</i> [118]	2019	Extract prognosticators	✗	✓	H&E images	VGG19, AlexNet, ResNet50, GoogLeNet	Private dataset: 86 slides	Accuracy= 0.94.		
Hasan <i>et al.</i> [119]	2019	Polyp classification	✗	✓	Endoscopic videos	Alexnet, ResNet50, VGG16, VGG19	Private dataset: 317 frames	Accuracy= 0.9959, Sensitivity= 0.9974 and Specificity= 0.9944 (VGG19)		
Graham <i>et al.</i> [120]	2019	Segmentation of glands and Classification for different grades of cancer Gland	✓	✓	H&E stained images	Minimal Information Loss Deep neural network (MILD-Net)	GlaS dataset: 165 images, CRAG dataset: 213 images	F1-score= 0.825, Dice= 0.875 (on GlaS), F1-score=, Dice= 0.867 (on CRAG)		
Wang <i>et al.</i> [79]	2021	Classification of normal, polyp, inflammation, and cancerous images	✗	✓	Colonoscopy	AlexNet, VGG16, VGG19, ResNet50, ResNet101	Private dataset: 2150 images.	Accuracy= 0.8644 (ResNet-101).		
Tsai and Tao [121]	2021	Classification of colon tissues	✗	✓	H&E images	AlexNet, SqueezeNet, VGG19, GoogLeNet, ResNet50	NCTCRC-HE-100K, CRC-VAL-HE-7K, Kather-texture-2016-image: 5000 images	Accuracy = 0.9932 (on CRC-VAL-HE-7K) Accuracy = 0.9969 (on NCTCRC-HE-100K) Accuracy= 0.9486 (on Kather-texture-2016-image) (by ResNet50)		
Jha <i>et al.</i> [43]	2021	Polyps segmentation	✓	✗	Colonoscopy images	NanoNet inspired from MobileNet and U-Net	Kvasir-SEG: 1000 images	Dice=0.8227, meanIoU= 0.7282, Accuracy= 0.9456		
Tomar <i>et al.</i> [122]	2021	Polyps segmentation	✓	✗	Colonoscopy images	DDANet	Kvasir-SEG: 1000 images	Dice= 0.8576, meanIOU= 0.78		
Hamida <i>et al.</i> [123]	2021	Colon tumor classification and segmentation	✓	✓	WSI	AlexNet, VGG, ResNet, DenseNet and Inception, U-Net, SegNet	AiCOLO dataset: 396 slides, CRC-5000: 5000 images, NCT-CRC-HE: 100,000 images	Accuracy=0.9677 (on CRC-5000), 0.9976 (on NCT-CRC-HE) by ResNet, Dice= 0.9838 (on CRC-500), and 0.9873 (on NCT-CRC-HE) by SegNet		
Sharma <i>et al.</i> [68]	2022	Polyps segmentation	✓	✗	Colonoscopy	Customized CNN with attention	Kvasir-SEG: 1000 images, CVC-ClinicDB: 612 images	meanIoU= 0.88 and dice= 0.9058 (on Kvasir-SEG), meanIoU= 0.8969 and Dice= 0.9372 (on CVC-ClinicDB)		
Huang et al. [124]	2022	Polyp segmentation	✓	✗	Colonoscopy	CNN with hybrid channel-spatial attention	Kvasir-SEG: 1000, CVC-ClinicDB: 612, CVC-ColonDB:380, ETIS-Larib PolypDB: 196, CVC300:300 iamges	Dice= 0.916 (Kvasir), 0.949 (CVC-ClinicDB), 0.80 (CVC-ColonDB), 0.911 (CVC300), 0.75 (ETIS)		
Liu <i>et al.</i> [125]	2023	Polyp segmentation	✓	✓	Colonoscopy	U-Net+ attention	Kvasir-SEG, CVC-300, CVC-ClinicDB, CVC-ColonDB and ETIS-LaribPolypDB	Dice= 0.9135 (on Kvasir-Seg)		
Wu et al. [126]	2023	Polyp segmentation	✓	✗	Colonoscopy	CNN with swine transformer and channel attention	Kvasir:1000 images, CVC-ClinicDB: 612 images	Dice= 0.929(Kvasir), 0.928(CVC-ClinicDB)		

U-Net, Tomar *et al.* [122] proposed a model named DDANet that takes the backbone of U-Net and introduces residual connection in the convolutional block of the original U-Net. Although this method achieved better results than the U-Net, it is computa-

Table 6: Summary of existing CNN models used for prostate cancer detection. S → Segmentation, C → Classification.

Authors	Year	Application	Task		Modality	CNN Architecture	Dataset	Performance
			S	C				
Wang <i>et al.</i> [127]	2017	Prostate cancer detection	✗	✓	MRI	CNN	Private dataset: 1164 images	AUC=0.84, Sensitivity= 0.696, Specificity= 0.839
Yoo <i>et al.</i> [128]	2019	Prostate cancer detection	✗	✓	MRI	Customized ResNet	Private dataset: 5832 images	DWI slice-level AUC= 0.87 and patient-level AUC= 0.84
Abbasi <i>et al.</i> [129]	2020	Classification of prostate and brachytherapy	✗	✓	MRI	GoogLeNet	Generated public dataset: 682 images	AUC= 1.0
Salvi <i>et al.</i> [130]	2021	Detect the contours of prostate glands	✓	✗	WSI	Hybrid U-Net with active contour	Generated public dataset: 1500 images	Precision= 0.9124 , Recall= 0.9723.
Vente <i>et al.</i> [131]	2021	Prostate zone segmentation and grading of prostate tissue	✓	✓	MRI	2D U-Net	ProstateX-2: 182 images	Dice= 0.370 ± 0.046
Chen <i>et al.</i> [132]	2021	Automatic segmentation of prostate images	✓	✗	MRI	3D AlexNet, ResNet-50, Inception-V4	Private dataset: 500 images	Accuracy= 0.921, Sensitivity= 0.896, Specificity= 0.902, F-score=0.897, AUC= 0.964 (3D AlexNet).
He <i>et al.</i> [133]	2021	Localization and segmentation of prostate region	✓	✗	CT	Customized U-Net	Private dataset: 339 patients data	Dice= 0.88, Sensitivity= 0.88
Hamarde <i>et al.</i> [134]	2021	Segmentation of prostate gland and prostate lesions	✓	✗	MRI	U-Net	Private dataset: 3245 images	For gland segmentation: Dice= 0.9921, and for lesion segmentation: Dice= 0.9965.
Silva-Rodriguez <i>et al.</i> [135]	2021	Detect local cancerous patterns in the prostate tissue	✓	✗	HistologyWeGleNet		Generated dataset by [136]: 886 images	F1-score= 0.58
Hoar <i>et al.</i> [137]	2021	Pixel-by-pixel prediction of cancer vs. non-cancerous prostate images	✓	✗	Multi-parametric MRI	Customized VG-Net	Private dataset: 135 volume	Dice= 0.59.
Duran <i>et al.</i> [69]	2022	Segmentation of the prostate gland and classification of cancer lesions GS group grade	✓	✓	mp-MRI	CNN with attention	Private dataset: 219 MRI, PROSTATEx-2: 182 lesions	Dice= 0.875 (for segmentation on PROSTATEx-2), Sensitivity= 0.69 (for classification on PROSTATEx-2).
Wang <i>et al.</i> [138]	2023	Presence of prostate in MRI and segmentation of prostate from background	✓	✓	MRI	Customized U-Net	Private dataset: 81 patient	Dice = 0.86 and IoU= 0.758
Ding <i>et al.</i> [139]	2023	Prostate segmentation	✓	✗	MRI	U-Net with self-attention and transformer	PROSTATEx dataset: 182 images	Dice= 0.9093 and IoU= 0.8345
Liu <i>et al.</i> [140]	2023	Segmentation whole tumor (WT), tumor core (TC), and enhancing tumor (ET)	✓	✗	MRI	U-Net with attention and dilation convolution	BraTS 2019: 335 images	Dice= 0.7811 (ET), 0.9002 (WT) 0.8415 (TC)

tionally expensive. As the colonoscopy process demands real-time analysis, the CAD systems require to be lightweight. In this line, Jha *et al.* [43] developed a lightweight model replacing the encoder with the popular MobileNet model and achieved better performance while reducing the parameters drastically. But, these models still need improvements in performance to be applied in clinical applications. Moreover, issue of polyp overlooked due to its small and flat characters demand more accurate and robust feature extraction methods. Fortunately, CNN with an embedded attention module has shown promising results in polyp segmentation and classification and

partially decreases the polyp miss-rate [68, 122, 125, 124, 126].

5.4. Prostate cancer

Prostate cancer is one of the most common cancers in men which grows very slowly. It is highly treatable if it can be diagnosed at early stages. Current clinical procedures for diagnosing prostate cancer include a blood test for prostate-specific antigen (PSA), digital rectal examination (DRE), transrectal ultrasonography (TRUS), needle biopsy, and magnetic resonance imaging (MRI). Due to PSA testing, the incidence of prostate cancer patients diagnosed at an earlier stage has grown during the last few years [143, 144]. However, misidentification of the diseased area may delay the treatment of the condition or cause over-diagnosis. MRI is a non-invasive and has been found to be an alternative for the accurate detection of prostate cancer. However, accurate interpretation of MRI images requires skilled supervision and it is very time-consuming. Thus, the design of automated methods using MRI is crucial for the early detection of prostate cancer. In recent years, in regard to automated prostate cancer diagnosis, CNNs are gaining popularity which are summarized in Table 6. From the reviewed articles, it has been seen that for the prostate cancer analysis, segmentation is the most focused area of research [130, 131, 132, 134, 135, 137], and U-Net is the most dominant architecture [131, 134, 133, 140]. He *et al.* [133] developed a customized U-Net and achieved a better result than the original U-Net. It is evident that the incorporation of attention mechanisms is a major contributor to the enhanced performance of the customised U-Net [139, 69, 140]. Chen *et al.* [132] used ResNet50, Inception and AlexNet and extended it from 2D to 3D architecture. Relatively a few articles focused on classification task [127, 128, 129, 138]. It is worth noticing that majority of the existing studies utilized private datasets which are not available publicly and therefore, the proposed trained models may not be useful in other hospitals.

5.5. Skin cancer

Dermatologists follow a series of processes to analyze skin disease. The process first begins with a naked-eye analysis, if any suspected abnormality is observed then dermatoscopy is the next process and this process is supported by a biopsy. Automated methods based on machine learning and deep learning using dermoscopic images have been developed in the past decade to aid dermatologists in clinical decision-making and to detect extremely suspicious cases of skin disease [145, 146, 148]. Abnormal growth of epithelial tissues of the skin form tumor. This disease primarily develops in the melanocytes cells, which produce melanin. Malignant skin tissues are called melanoma. Basal cell carcinoma and Squamous cell carcinoma are two types of non-melanoma cells. Malignant melanoma is the most lethal one among all other skin cancers and the only reason for 77% deaths due to skin cancer. Early diagnosis is essential because the survival rate is only 14% if diagnosed later in the disease, whereas the survival rate is more than 90% if diagnosed early. Several CNN architectures have been explored to detect skin cancer due to the availability of a few

Table 7: Summary of existing CNN models used for skin cancer detection. S → Segmentation, C → Classification.

Authors	Year	Application	Task		Modality	CNN Architecture	Dataset	Performance
			S	C				
Yuan <i>et al.</i> [145]	2017	Melanoma skin lesion segmentation	✓	✗	Dermoscopic image	Customized CNN	ISIC 2016 , ISBI 2016: 1279 images, PH2:200 images	Dice= 0.912 (on ISBI 2016) Dice= 0.938 (on PH2 dataset).
Kassani and Kassani [146]	2019	Melanoma detection	✗	✓	Dermoscopic image	ResNet, VGG, AlexNet, Xception	ISIC 2018: 10,015 images	Accuracy= 0.9208, F1-score= 0.9274. (by ResNet50).
Saba <i>et al.</i> [147]	2019	Boundary detection of skin lesions	✓	✗	Dermoscopic images	Customized Inception V3	ISBI 2016: 1279 images, ISBI 2017: 2750 images, PH2: 200 images, ISIC 2017: 2000 images	Accuracy (for classification)= 0.984 (PH2), 0.951 (ISBI 2016), and 0.948 (ISBI 2017), Accuracy (for segmentation)= 0.9541 (PH2) and 0.9478(ISIC 2017)
Wei <i>et al.</i> [148]	2020	Melanoma and non-melanoma cell recognition	✓	✓	Dermoscopic images	MobileNet, DenseNet and U-Net	ISBI 2016:900 images	Accuracy= 0.962, Dice= 0.923 (for segmentation), and Accuracy= 0.876, Jaccard= 0.87, AUC= 0.854 (for classification)
Ashraf <i>et al.</i> [149]	2020	Melanoma detection	✗	✓	Dermoscopic images	Customized CNN	DermIS: 206 images and DermQuest	Recall= 0.968, F1-score= 0.961 (on DermIS dataset), and Recall= 0.769, F1-score= 0.76 (on DermQuest)
Kadampur <i>et al.</i> [150]	2020	Skin cancer prediction	✗	✓	Dermoscopic cell images	ResNet, SqueezeNet, DenseNet, InceptionV3	HAM10000:10,015 images	Precision= 0.9819, Recall= 0.9574 and AUC= 99.23 (by InceptionV3)
Guergueb and Akhloufi [151]	2021	Benign and malignant image classification	✗	✓	Deremoscopic images	VGG, DenseNet, ResNet, EfficientNet, MobileNet, InceptionResNetV2, Xception and Inception	Combined dataset of ISIC 2020, ISIC 2019, ISIC 2018 and ISIC 2017: 36,502 images	Accuracy = 0.9933 (EfficientNet)
Shorfuzzaman [152]	2021	Malignant and benign melanoma detection	✓	✗	Dermoscopic images	Ensemble of DenseNet, Xception and EfficientNetB0	ISIC dataset:2750images	Accuracy= 0.9576, F1-score= 0.9613, AUC= 0.957
Ali <i>et al.</i> [153]	2021	Classification of benign and malignant skin lesions	✗	✓	Dermoscopic images	AlexNet, ResNet, VGG-16, DenseNet, MobileNet	HAM 10000: 10015 images	Accuracy= 0.9193
Bagheri <i>et al.</i> [154]	2021	Skin lesion segmentation	✓	✗	Dermoscopic images	CNN	ISBI 2016: 1279 images, ISBI 2017: 2750 images, ISBI 2018: 2594 images, PH2: 200 images	Jaccard= 0.8983 (PH2), 0.8721 (ISBI 2016), 0.7997 (ISBI 2017)
Reis <i>et al.</i> [155]	2022	Benign and malignant lesions classification	✗	✓	Dermoscopic images	Customized CNN	HAM10000 images, ISIC 2019, and ISIC 2020	Accuracy= 0.9459 (in ISIC 2018), 0.9189 (in ISIC 2019), and 0.9054 (in ISIC 2020).
Gajera <i>et al.</i> [156]	2022	Melanoma classification	✗	✓	Dermoscopic images	EfficientNetB0	ISIC 2016	Accuracy= 0.8707
Hu <i>et al.</i> [157]	2022	Skin lesion segmentation	✓	✗	Dermoscopic images	VGG with spatial and channel attention	PH2: 200 images, ISIC 2018: 2594 images	Accuracy= 0.952 and Dice= 0.9305 (on PH2) and Accuracy= 0.9568 and Dice= 0.8955 (on ISIC 2018)
Kumar <i>et al.</i> [158]	2023	Segmentation and multi-class skin cancer classification	✓	✓	Dermoscopic images	End-to-end CNN	ISIC dataset: 2357 images	Accuracy= 0.9746, Precision= 0.93627, Recall= 0.9972, and F1-score= 0.9718.

publicly available datasets such as ISBI 2016, ISBI 2017, and HAM 10000. The reviewed articles as shown in Table 7 reveal that both segmentation and classification tasks have been investigated frequently, and most of these studies either used pre-trained CNN models or proposed customized CNN models. Among the pre-trained networks, ResNet, VGG, AlexNet, MobileNet and EfficientNet are most popular models [146, 150, 151, 153, 156]. The inception model was used by Kadarmour and Al Riyae [150] to extract features in multiple scales from the dermoscopic images. Kassani and Kassani [146] and Guergueb and Akhloufi [151] used Xception model, which is light-weight compared to Inception. Recently, Shorfuzzaman [152] used an ensemble approach to improve the performance of a single pre-trained model. However, due to the higher computational cost of these pre-trained models, many studies proposed

Table 8: Summary of existing CNN based liver cancer detection methods. S → Segmentation, C → Classification.

Authors	Year	Application	Task		Modality	CNN Architecture	Dataset	Performance
			S	C				
Vivanti <i>et al.</i> [159]	2017	Existing tumor localization and segmentation, and new tumor segmentation	✓	✗	CT	CNN	Private dataset: longitudinal liver CT	Recall= 0.86 (new tumors) and 0.72 (existing tumor).
Trivizakis <i>et al.</i> [160]	2019	Liver tissue classification	✗	✓	Diffusion weighted MRI (DW-MRI)	3D-CNN	Private dataset: 130 patients	Accuracy= 0.83(3D-CNN), and 0.652 (2D-CNN).
Das <i>et al.</i> [161]	2019	Classification of liver cancer into three groups	✗	✓	CT	CNN	Private dataset: 225 images	Accuracy= 0.9938, Jaccard index= 0.9818.
Ozyurt <i>et al.</i> [162]	2019	Classification of benign and malignant masses	✗	✓	CT	F-PH-CNN	Private dataset: 75 images	Accuracy= 0.946.
Zhang <i>et al.</i> [163]	2020	Segmentation of liver tumor	✓	✗	CT	3D U-Net	Liver tumor segmentation challenge (LiTS): 201 images	Dice= 0.953.
Araujo <i>et al.</i> [164]	2021	Automatic segmentation of liver lesion	✓	✗	CT	RetinaNet, U-Net	LiTS: 131 CT images	Sensitivity= 0.8386, Specificity= 0.9996, and Dice= 0.8299.
Lal <i>et al.</i> [165]	2021	Automatic nuclei segmentation of liver cancer	✓	✗	HistopathologyNucleiSegNet	KMC liver dataset: 80 images, KUMAR dataset: 44 images	F1-score= 0.81363 Jaccard= 0.68883 (on KUMAR dataset), F1-score= 0.8359 Jaccard= 0.7206 (on KMC dataset)	
Chi <i>et al.</i> [166]	2021	Segment liver and tumor	✓	✗	CT	XNet: Customized U-Net	MICCAI 2017 LiTS: 201 CT volumes, 3DIRCADb: 20 CT volumes	For LiTS dataset: Dice = 0.971 (for liver), Dice= 0.843 (for tumor), for 3DIRCADb dataset: Dice= 0.9668 (for liver), 0.6911 (for tumor)
Gao and Almekkawy [167]	2021	Liver tumor segmentation	✓	✗	Ultrasound, CT	U-Net++	SYSU-US: 480 images, SYSU-CT: 135 images, and subCT: 118 images	Dice= 92.35 IOU= 85.34 (on SYSU-US), Dice= 0.9387 IOU= 0.8847 (on SYSU-CT) and Dice= 0.9205 IOU= 0.8527 (on subCT)
Aghamohammadi <i>et al.</i> [168]	2021	Distinguish border for liver and tumour	✓	✗	CT	Cascade CNN	Generated dataset by [169]: 20,000 slices	RMS= 2.3±0.1 (for Liver segmentation), and RMS= 3.7±0.4 (for tumor segmentation)
Liet <i>et al.</i> [70]	2022	Liver and liver tumor segmentation	✓	✗	CT	Improved U-Net++	LiTS dataset: 201 images	Dice= 0.958 (for liver segmentation) and 0.893 (for liver tumor segmentation)
Zhang <i>et al.</i> [170]	2022	Tumor segmentation	✓	✗	CT	U-Net with dynamic scale attention	LiTS2017: 131 images	Dice= 0.8449
Rela <i>et al.</i> [171]	2023	Liver tumor segmentation and classification	✓	✓	CT	U-Net and customized CNN	Combined dataset of LiTS dataset and Private dataset: 206 images	Accuracy= 0.95128, Dice= 0.9747, F1-score= 0.9747.
Li <i>et al.</i> [70]	2023	Liver and Liver tumor segmentation	✓	✗	CT	U-Net++ with channel attention	LiTS dataset: 200 images	Accuracy= 0.958 and 0.893 (Liver tumor segmentation)

to use customized CNN architectures [145, 147, 149, 154, 156] and tried to minimize the parameters used in the model without compromising the performance. In [145] and [154], authors designed customized CNN for the segmentation of the skin lesions. While Wei *et al.* [148] used U-Net for the segmentation task.

5.6. Liver cancer

The liver is an essential organ for human health that filters toxic substances and generates digestive biochemicals. For liver cancer detection and proper treatment planning with the aid of an automated system, segmentation and classifica-

Table 9: Summary of existing CNN models used for brain tumor detection. S → Segmentation, C → Classification.

Authors	Year	Application	Task		Modality	CNN Architecture	Dataset	Performance
			S	C				
Deepak and Ameer [172]	2019	Brain tumor classification	✗	✓	MRI	Modified GoogLeNet	Figshare dataset: 3064 iamges	Accuracy= 0.93
Ismael <i>et al.</i> [173]	2019	Brain tumor classification	✗	✓	MRI	ResNet50	Figshare dataset: 3064 iamges	Accuracy= 0.97
Naser <i>et al.</i> [174]	2020	Brain tumor segmentation and grading	✓	✓	MRI	Customized VGG16 and UNet	TCIA: 7858 images	Accuracy= 0.92, (segmentation), and Accuracy= 0.89 (grading)
Mehrotra <i>et al.</i> [175]	2020	Malignant and benign brain tumor classification	✗	✓	MRI	AlexNet, GoogLeNet, ResNet-50, ResNet-101, SqueezeNet	TCIA: 696 images	Accuracy= 0.99 (AlexNet)
Tandel <i>et al.</i> [176]	2021	Brain tumour grading	✗	✓	MRI	AlexNet, VGG16, ResNet18, GoogleNet, and ResNet50	REMBRANDT: 2908 images	Accuracy= 0.9677, Sensitivity= 0.9438, Specificity= 0.9789, and AUC= 0.9614
Karayegen <i>et al.</i> [177]	2021	Brain tumor segmentation	✓	✗	MRI	Customized DNN	BraTS: 257 images	Accuracy= 0.957, F1-score= 0.93, IoU= 0.8812
Wang <i>et al.</i> [178]	2021	Brain tumor segmentation	✓	✗	MRI	CLCU-Net: customized U-Net	BraTS 2018: 351 images	Dice= 0.885, Precision= 0.9198, Recall= 0.8562
Sharif <i>et al.</i> [179]	2021	Detection of high-grade glioma (HGG) and low-grade glioma (LGG)	✗	✓	MRI	InceptionV3	BraTS2013: 30 images, BraTS2014: 300 images, BraTS2017: 285 images and BraTS2018: 351 images	Accuracy= 0.983 (BraTS2013), 0.978 (BraTS2014), 0.969 (BraTS2017), and 0.925 (BraTS2018)
Zhou <i>et al.</i> [180]	2021	Brain tumor segmentation	✓	✗	MRI	ERV-Net: 3D residual neural network	BraTS 2018: 351 images	Dice= 0.9121 (for whole tumor) and 0.8662 (for tumor core)
Huang <i>et al.</i> [181]	2021	Whole brain tumor (WBT), tumor core (BTC) and enhancing brain tumor (EBT) segmentation	✓	✗	MRI	BraTS 2017: 285 images, BraTS 2018: 351 images	U-Net with cross channel attention	Dice (on BraTS 2017)= 0.872 (WBT); 0.796(BTC); 0.781(EBT), Dice (on BraTS 2018)= 0.909 (WBT); 0.845(BTC); 0.813(EBT)
Hou <i>et al.</i> [182]	2021	Classification of Glioblastoma Tumor into four class (normal tissue, tumor tissue, hypervasculated tissue, and background)	✗	✓	HSI	Fused CNN	Private dataset: HSI 36	Accuracy: 0.9669, Precision: 0.9268, Recall: 0.9082
Ottom <i>et al.</i> [183]	2022	Brain tumor segmentation	✓	✗	MRI	ZNet	Cancer Genome Atlas Low Grade Glioma (TCGA-LGG): 3,929 slices	Dice: 0.9158, F1-score: 0.8098
Dutta and Nayak [64]	2022	Brain tumor classification	✗	✓	MRI	CDANet	Figshare dataset: 3064 iamges	Accuracy= 0.966
Yousaf <i>et al.</i> [184]	2023	Brain tumor segmentation	✓	✗	MRI	Improved U-Net	Merged BRATS 2015 and ISLES 2015: 448 images	Accuracy= 0.9956, Specificity= 0.9999, and F1-score= 0.9957

tion of liver and liver vascular are very essential. However, a large variety of tissue appearance makes the segmentation and classification of liver tumors a complex task [160]. From the reviewed articles on automatic liver cancer analysis, we found that the majority of the work focused on segmentation tasks and utilized CT scan images [159, 161, 163, 164, 70, 171, 170]. Some recent studies on liver cancer detection using CNN are summarized in Table 8. U-Net is the most widely used CNN-based architecture for segmentation. Its structure is not only efficient for the segmentation of

tumors, but also highly efficient in analyzing histopathology images [165]. Although standard U-Net achieves satisfactory performance, to make it more adaptable to the complex structure of the liver tissues, Chi *et al.* [166] and Zhang *et al.* [163] modified the U-Net architecture for liver and tumor segmentation from CT volumes. In a few studies, different CNN-based encoder-decoder architectures customized with residual and inception blocks have been proposed for segmentation of liver CT and histology images [159, 161, 164, 168, 167, 171]. Further, limited efforts have been made for solving classification tasks. In general, grading of tumors is essential after segmentation of the tumor regions to design a complete CAD system. Das *et al.* [161] focused on classifying the tumor into three groups using 3D-CNN. The classification of benign and malignant liver tissues was performed in [160] and [162].

5.7. Brain cancer

A brain tumor is a mass-like structure of alive and dead cells that start to grow uncontrollably inside the brain. It is considered one of the deadliest forms of cancer across the globe that affects individuals of all age groups, including children [185]. Hence, early detection of brain tumors is highly essential. Plenty of automated methods based on machine learning and deep learning have been proposed in the literature for brain tumor detection; however, it is still an open problem due to the inter-class similarity among the different varieties of brain tumors. A summary of CNN-based recently published methods for brain tumor diagnosis is presented in Table 9. From the reviewed articles, it can be observed that MRI is the most widely used imaging modality for brain cancer treatment [172, 174, 176, 178, 180, 183, 64]. A number of studies utilized pre-trained CNN models to perform brain tumor classification [172, 176, 175, 174]. Recently, a channel split dual attention-based CNN named CDANet [64] was proposed to capture the prominent and more focused features from the brain MR images. Segmentation of brain tumors is one of the most popular task in the domain of medical image analysis. Several attempts have been made in the past decade to achieve human-level performance. Recently, CNN has shown improved performance over the traditional segmentation methods. For instance, Naser *et al.* [174] used U-Net for the segmentation of brain tumors, while Karayegen *et al.* [177] and Ottom *et al.* [183] proposed a customized DNN and ZNet architecture, respectively, for segmenting brain tumors. In [184], an improved U-Net model was proposed to segment brain tumors. In another contribution, Huang *et al.* [181] proposed a segmentation model based on U-Net aided with group cross-channel attention and achieved promising performance gain compared to the original U-Net. Most of the articles reviewed in this study used publicly available datasets [173, 174, 175, 177, 183]. Also, it has been seen that few studies utilized private datasets to validate their models [186, 182].

6. Datasets

Knowing the importance of supervised learning techniques for automated disease detection, a group of researchers and organizations have developed various datasets by collecting medical images from radiology centers, hospitals, and cancer research institutes. Many datasets in the reviewed papers have been kept private due to privacy concerns, and reproducing their results is hence not possible. Table 10 lists the most widely used publicly available datasets related to each cancer type reviewed in this survey and their descriptions, including the main reference, imaging technique, number of images, and the URL. It can be seen that there are no large-scale datasets available to date in the medical imaging field like ImageNet, which often poses challenges in designing effective CNN models.

Table 10: List of publicly available datasets

Cancerous Area	Datasets	Image Type	Quantity
Lung	LIDC-IDRI [187]	CT	224,527 images
	ANODE09	CT	55 scans
	DLCST	CT	4101
	JSRT	X-ray	647
	ACDC-Lung HP	Histopathology	200 images
	ELCAP [188]	CT	50 scan
	NSCLC	CT	52073 images
Breast	MITOS-ATYPIA	Histology	1620 frames
	INbreast [189]	Mammograms	410 images
	BreakHis	Microscopic images	7,909 images
	MIAS/ miniMIAS [190]	Mammography	330 images
	DDSM [191]	Mammography	10480 images
	ICPR 2014 MITOSIS [192]	Histopathology slides	1420
Colon	Kvasir [193]	Endoscopy	8000 images
	Kvasir-SEG [194]	Endoscopic	1000 images
	Nerthus Data-set [195]	Endoscopy Video	21 videos with 5, 525 frames
	ASU-Mayo Clinical Database [196]	Colonoscopy video	38 video
	CVC-ClinicDB	colonoscopy	31 sequences
	CRCHistoPhenotypes	Histology Images	100 images
	EndoTech 2020	Colonoscopy frames	1000
Prostate	Hyper-kvasir	Images with binary mask	1000
	PROSTATE-MRI TCIA [197]	MRI	22,036 images
	PROSTATEx Challenges	MRI	538 images
	PROSTATEx-2 Challenges	MRI	162 images
	PROMISE12	MRI	70
Skin	DermIS	Dermoscopic image	206 images
	DermQuest	Dermoscopic image	
	ISIC	Dermatology images	2750
	ISBI 2016	Dermoscopic images	1279
	HAM 10000 [198]	Dermatoscopic images	10015 images
	MED-NODE	Macroscopic images	170 images
Liver	IRCAD [199]	CT	20 patient with varying number of slides
	MIDAS [200]	CT	30 volume
	MICCAI Sliver'07	CT	30 patient data
	PAIP 2019 [201]	WSI	100 images
Brain	IBSR	MRI	38 scans
	OASIS	Multi-model	MRI:2168, PET:1608
	BRATS (2012-2019)	MRI	65 images
	FIGSHARE CJDATA	MRI	3064 images

7. Challenges and future directions

From the reviewed papers, it is found that there has been a tremendous stride in analyzing medical images using CNN architectures. The performance of state-

of-the-art CNN models on medical images is not as good as natural images. This is because the medical images are highly complex in structure, and gray-scale differences between different tissue classes of the human body are significantly less. Also, it is observed that there has been a profound improvement in early cancer detection via the CNN approach compared to earlier detection methods due to its ability in learning high-level features from medical images, but it is still an open problem due to certain challenges and issues.

7.1. Emerging challenges and future directions

In this section, we provide a list of challenges that require significant attention from the research community. Also, we enumerate potential research directions that can be explored to effectively address the challenges in the future. The visual representation of the challenges and future direction is illustrated Figure 8.

7.1.1. Lack of large training datasets

The lack of large training datasets is one of the biggest challenges researchers face in training CNN models to analyze medical images. Even though the technological improvements help hospitals store patient data (images), it is not easily accessible for research since most of them contain confidential information about patients. From the surveyed papers, it has been noticed that many papers used private datasets collected from hospitals or cancer research institutes to validate the CNN models; however, reproducing similar results and comparison with such models has become challenging. Although several publicly accessible datasets have been available recently, most are smaller in size. And, training CNN models on these datasets will lead to overfitting problems. Hence, there is a strong need to develop large and diverse open source datasets for different cancer-related tasks. Data augmentation has been adopted in a few surveyed papers to combat the above limitations. The major data augmentation techniques include different image transformations such as cropping, shearing, rotation, mirroring, skewing, flip, scaling, adding noise, etc., which are used to increase the size of a given datasets [12, 80, 96, 85]. Transfer learning has been found to be an alternative solution to avoid the limitations of large datasets [12], thereby reducing the chance of overfitting. Hence, among all surveyed papers, an ample amount of papers adopted transfer learning for various cancer detection tasks [107, 12, 89, 118, 119, 123, 129, 146, 148, 175]. In a few of them, pre-trained CNN models were utilized as the feature extractor, whereas the pre-trained CNN models were fine-tuned on medical image datasets in other studies. Also, it is seen that the state-of-the-art CNN models were pre-trained using a large-scale dataset known as ImageNet, which does not contain medical images. And, there is no such CNN model yet available which is pre-trained on medical image data. Hence, it is expected to develop CNN models on large-scale medical image datasets and use them via transfer learning in future studies. The design of lightweight CNN models has recently attracted considerable attention to avoid the issue of overfitting [202].



Figure 8: Visual depiction of challenges and future directions in employing CNN for cancer detection from medical images.

Further, in recent years, advanced CNN models like generative adversarial networks (GANs) have been investigated to generate high-resolution medical images, thereby overcoming the problem of limited training data [203, 204, 205]. Additionally, the use of self-supervised learning mechanisms have recently been explored in a few studies to handle the issue of limited labelled training data [113].

7.1.2. Issue of data annotation

The collection and storage of medical images have become a relatively easier task, whereas annotating or labeling such images has turned out to be the most challenging job. Usually, annotations are made by the domain experts such as radiologists, pathologists, etc., for a given task. However, labeling each image in a large dataset is time-consuming and troublesome. For example, to segment brain tumors using CNN models, radiologists need to annotate the MRI data slice by slice, which is te-

dious. Therefore, it is essential to design CNN models which can learn effectively from limited annotated data. Also, future studies include the use of semi-supervised algorithms in the domain of medical image analysis to automatically label unlabelled images.

7.1.3. Issue of Data Imbalance

Data imbalance is another big concern in this field of research. This is due to the difficulty in collecting or finding images for the abnormal class compared to the normal or healthy class for some specific task. For example, consider a dataset ISBI 2016 with 1031 non-melanoma and 248 melanoma cases, which has been extensively used for skin cancer detection. The problem of data imbalance causes the CNN model to be biased towards the more common class (e.g., non-melanoma in the above example). To overcome the above issues, data augmentation has been adopted widely [204]. However, the development of CNN models that can handle this issue is an open research problem.

7.1.4. Usage of multi-modal data

Recent years have witnessed the potential of using multi-modal data for arriving at better clinical decisions. For example, different MR modalities such as T1-weighted, T2-weighted, and Fluid-Attenuated Inversion Recovery (FLAIR) have been used to detect brain tumors accurately [206]. However, handling multi-modal data using CNN is relatively challenging. Therefore, designing effective CNN models to handle multi-modal image data is an important area of research. Further, the importance of clinical information along with the image information has been investigated in many recent studies [207, 208, 209, 210] and it is envisaged to see many more such studies in near future.

7.1.5. Issues with computationally complex convolutional neural networks models

The use of more computationally complex CNN architectures has become burdensome due to resource constraints, and requirements of low latency and fast inference in real time clinical applications. Additionally, IoT devices require lightweight CNNs to enable intelligent processing and decision-making at the edged devices. Therefore, developing lightweight CNNs are crucial to increase the clinical applicability of CNN in real-time. To this end, recently researchers have been trying to develop CNN models with comparatively less number of learnable parameters while maintaining the computational complexity and precise decision [43, 211, 212, 213].

7.1.6. Usage of transformer-based deep learning models

Recently, ViT models have achieved state-of-the-art performance in a wide range of computer vision tasks due to their ability to capture global feature inter-dependencies with the use of self-attention mechanism [75]. Inspired by their success, the medical imaging research community has shown a growing interest in applying ViT models for cancer detection through medical images [214, 215]. However, the demand of a

large number of parameters and large-scale datasets to achieve desirable performance makes ViT models challenging to use in real-time applications in medical images. Therefore, efforts toward development of efficient ViT based models need to be made in analyzing medical images.

7.1.7. Lack of explainability

Traditional CNN acts as a black box, implies that it does not provide clear explanation about the reason behind its predictions or decisions. This lack of interpretability makes hard to trust and effectively utilize CNN in clinical applications. To address this issue, Explainable AI, an approach where feature visualization, feature importance and saliency analysis, model-agnostic interpretability are adopted, providing interpretable and transparent insights on the decision made by a CNN model. On cancer diagnosis from medical images, a scanty amount of studies have yet been explored the concept of explainable AI [216, 217, 218] Therefore, a plethora of research work on the explainability of CNN models are expected in near future, which can improve the trust and ultimately encourage healthcare professionals in using CNN-based systems for clinical applications.

In a nutshell, CNNs have strongly impacted the medical imaging research community and have obtained state-of-the-art results in many cancer detection tasks. However, it is worth noting that the application of CNN models in real practice poses several problems such as (1) They often demand a large-scale dataset for desirable performance and (2) The training of CNN models is computationally more costly and requires GPU-based hardware (3) Although many approaches have recently been proposed for understanding the deep structure of CNN and its layers, it is often regarded as black boxes. Hence, a suitable approach for this is the need of the hour, which would increase the acceptance of CNN models in hospitals and medical research institutes.

8. Concluding remarks

In this paper, we provided a comprehensive overview of current CNN-based methods employed for detecting cancer from various medical images, focusing on the novel aims of automating the diagnostic process to lower the mortality rate and reduce the workload of physicians. The review majorly reported the classification and segmentation tasks on seven cancer types: breast, lung, colorectal, prostate, liver, skin, and brain. Also, we highlighted various CNN architectures and their usage in this survey. Further, a separate discussion on datasets and a list of a few open-source datasets were provided, which can serve as a reference point for future research on medical image-based automated cancer diagnosis systems. Finally, we identified the crucial issues that emerged during the development of automated cancer detection through medical images and discussed a few potential solutions to combat such issues. This review would serve as a valuable reference for researchers and practitioners working on the classification and segmentation of medical images through CNN and modern

DL architectures. Moreover, this study will assist beginners who plan to contribute to this fascinating research domain.

References

- [1] W. H. Organization, Global health estimates 2015: deaths by cause, age, sex, by country and by region, 2000–2015, Geneva: WHO (2016).
- [2] R. L. Siegel, K. D. Miller, H. E. Fuchs, A. Jemal, et al., Cancer statistics, 2021, *Ca Cancer J Clin* 71 (1) (2021) 7–33.
- [3] F. Bray, M. Laversanne, E. Weiderpass, I. Soerjomataram, The ever-increasing importance of cancer as a leading cause of premature death worldwide, *Cancer* 127 (16) (2021) 3029–3030.
- [4] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA: a cancer journal for clinicians* 71 (3) (2021) 209–249.
- [5] X. Yang, L. Yu, S. Li, H. Wen, D. Luo, C. Bian, J. Qin, D. Ni, P.-A. Heng, Towards automated semantic segmentation in prenatal volumetric ultrasound, *IEEE Transactions on Medical Imaging* 38 (1) (2018) 180–193.
- [6] J. Chmelik, R. Jakubicek, P. Walek, J. Jan, P. Ourednicek, L. Lambert, E. Amadori, G. Gavelli, Deep convolutional neural network-based segmentation and classification of difficult to define metastatic spinal lesions in 3d ct data, *Medical Image Analysis* 49 (2018) 76–88.
- [7] T. Shen, Y. Wang, Medical image segmentation based on improved watershed algorithm, in: 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), IEEE, 2018, pp. 1695–1698.
- [8] R. Guetari, H. Ayari, H. Sakly, Computer-aided diagnosis systems: a comparative study of classical machine learning versus deep learning-based approaches, *Knowledge and Information Systems* (2023) 1–41.
- [9] K. Doi, Computer-aided diagnosis in medical imaging: historical review, current status and future potential, *Computerized Medical Imaging and Graphics* 31 (4-5) (2007) 198–211.
- [10] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R. M. Summers, Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning, *IEEE Transactions on Medical Imaging* 35 (5) (2016) 1285–1298.

- [11] T. Dhar, N. Dey, S. Borra, R. S. Sherratt, Challenges of deep learning in medical image analysis—improving explainability and trust, *IEEE Transactions on Technology and Society* 4 (1) (2023) 68–75.
- [12] P. Xi, C. Shu, R. Goubran, Abnormality detection in mammography using deep convolutional neural networks, in: 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA), IEEE, 2018, pp. 1–6.
- [13] T. A. Soomro, L. Zheng, A. J. Afifi, A. Ali, S. Soomro, M. Yin, J. Gao, Image segmentation for mr brain tumor detection using machine learning: A review, *IEEE Reviews in Biomedical Engineering* (2022).
- [14] E. H. Houssein, M. M. Emam, A. A. Ali, P. N. Suganthan, Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review, *Expert Systems with Applications* 167 (2021) 114161.
- [15] A. Mittal, S. Dhalla, S. Gupta, A. Gupta, Automated analysis of blood smear images for leukemia detection: a comprehensive review, *ACM Computing Surveys (CSUR)* 54 (11s) (2022) 1–37.
- [16] H. K. Gajera, D. R. Nayak, M. A. Zaveri, A comprehensive analysis of dermoscopy images for melanoma detection via deep cnn features, *Biomedical Signal Processing and Control* 79 (2023) 104186.
- [17] L. F. Sanchez-Peralta, L. Bote-Curiel, A. Picon, F. M. Sanchez-Margallo, J. B. Pagador, Deep learning to find colorectal polyps in colonoscopy: A systematic literature review, *Artificial intelligence in medicine* 108 (2020) 101923.
- [18] J. Wang, H. Zhu, S.-H. Wang, Y.-D. Zhang, A review of deep learning on medical image analysis, *Mobile Networks and Applications* 26 (2021) 351–380.
- [19] S. Maurya, S. Tiwari, M. C. Mothukuri, C. M. Tangeda, R. N. S. Nandigam, D. C. Addagiri, A review on recent developments in cancer detection using machine learning and deep learning models, *Biomedical Signal Processing and Control* 80 (2023) 104398.
- [20] Z. Hu, J. Tang, Z. Wang, K. Zhang, L. Zhang, Q. Sun, Deep learning for image-based cancer detection and diagnosis- a survey, *Pattern Recognition* 83 (2018) 134–149.
- [21] H. Yu, L. T. Yang, Q. Zhang, D. Armstrong, M. J. Deen, Convolutional neural networks for medical image analysis: state-of-the-art, comparisons, improvement and perspectives, *Neurocomputing* 444 (2021) 92–110.
- [22] B. Mughal, M. Sharif, N. Muhammad, T. Saba, A novel classification scheme to decline the mortality rate among women due to breast tumor, *Microscopy research and technique* 81 (2) (2018) 171–180.

- [23] W. Yang, L. Zhong, Y. Chen, L. Lin, Z. Lu, S. Liu, Y. Wu, Q. Feng, W. Chen, Predicting ct image from mri data through feature matching with learned non-linear local descriptors, *IEEE Transactions on Medical Imaging* 37 (4) (2018) 977–987.
- [24] M. Firmino, G. Angelo, H. Morais, M. R. Dantas, R. Valentim, Computer-aided detection (cade) and diagnosis (cadx) system for lung cancer with likelihood of malignancy, *Biomedical engineering online* 15 (1) (2016) 2.
- [25] V. Perrot, S. Salles, D. Vray, H. Liebgott, Video magnification applied in ultrasound, *IEEE Transactions on Biomedical Engineering* 66 (1) (2018) 283–288.
- [26] C. A. Santos, N. D. Mascarenhas, Geodesic distances in probabilistic spaces for patch-based ultrasound image processing, *IEEE Transactions on Image Processing* 28 (1) (2018) 216–226.
- [27] D. R. Nayak, R. Dash, B. Majhi, V. Prasad, Automated pathological brain detection system: A fast discrete curvelet transform and probabilistic neural network based approach, *Expert Systems with Applications* 88 (2017) 152–164.
- [28] E. D. Pisano, S. Zong, B. M. Hemminger, M. DeLuca, R. E. Johnston, K. Muller, M. P. Braeuning, S. M. Pizer, Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms, *Journal of Digital Imaging* 11 (4) (1998) 193.
- [29] D. R. Nayak, R. Dash, X. Chang, B. Majhi, S. Bakshi, Automated diagnosis of pathological brain using fast curvelet entropy features, *IEEE Transactions on Sustainable Computing* 5 (3) (2020) 416–427.
- [30] A. S. Panayides, A. Amini, N. D. Filipovic, A. Sharma, S. A. Tsafaris, A. Young, D. Foran, N. Do, S. Golemati, T. Kurc, et al., Ai in medical imaging informatics: current challenges and future directions, *IEEE Journal of Biomedical and Health Informatics* 24 (7) (2020) 1837–1857.
- [31] K. Kaplan, Y. Kaya, M. Kuncan, H. M. Ertunç, Brain tumor classification using modified local binary patterns (lbp) feature extraction methods, *Medical hypotheses* 139 (2020) 109696.
- [32] I. J. Hussein, M. A. Burhanuddin, M. A. Mohammed, N. Benameur, M. S. Maashi, M. S. Maashi, Fully-automatic identification of gynaecological abnormality using a new adaptive frequency filter and histogram of oriented gradients (hog), *Expert Systems* 39 (3) (2022) e12789.
- [33] S. Mourya, S. Kant, P. Kumar, A. Gupta, R. Gupta, Leukonet: Dct-based cnn architecture for the classification of normal versus leukemic blasts in b-all cancer, *arXiv preprint arXiv:1810.07961* (2018).

- [34] H. Irshad, S. Jalali, L. Roux, D. Racoceanu, L. J. Hwee, G. Le Naour, F. Capron, Automated mitosis detection using texture, sift features and hmax biologically inspired approach, *Journal of pathology informatics* 4 (2) (2013) 12.
- [35] S. Beura, B. Majhi, R. Dash, Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer, *Neurocomputing* 154 (2015) 1–14.
- [36] D. R. Nayak, R. Dash, B. Majhi, U. R. Acharya, Application of fast curvelet tsallis entropy and kernel random vector functional link network for automated detection of multiclass brain abnormalities, *Computerized Medical Imaging and Graphics* 77 (2019) 101656.
- [37] Y. Zhang, Z. Dong, L. Wu, S. Wang, A hybrid method for mri brain image classification, *Expert Systems with Applications* 38 (8) (2011) 10049–10053.
- [38] D. R. Nayak, Y. Zhang, D. S. Das, S. Panda, Mjaya-elm: A jaya algorithm with mutation and extreme learning machine based approach for sensorineural hearing loss detection, *Applied Soft Computing* 83 (2019) 105626.
- [39] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, C. I. Sánchez, A survey on deep learning in medical image analysis, *Medical Image Analysis* 42 (2017) 60–88.
- [40] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *nature* 521 (7553) (2015) 436.
- [41] V. Nair, G. E. Hinton, Rectified linear units improve restricted boltzmann machines, in: *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.
- [42] P. Sharma, A. Gautam, P. Maji, R. B. Pachori, B. K. Balabantaray, Li-segnet: Encoder-decoder mode lightweight segmentation network for colorectal polyps analysis, *IEEE Transactions on Biomedical Engineering* (2022).
- [43] D. Jha, N. K. Tomar, S. Ali, M. A. Riegler, H. D. Johansen, D. Johansen, T. de Lange, P. Halvorsen, Nanonet: Real-time polyp segmentation in video capsule endoscopy and colonoscopy, *arXiv preprint arXiv:2104.11138* (2021).
- [44] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556* (2014).
- [45] M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in: *International Conference on Machine Learning*, PMLR, 2019, pp. 6105–6114.

- [46] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2921–2929.
- [47] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
- [48] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.
- [49] G. Huang, Z. Liu, G. Pleiss, L. Van Der Maaten, K. Weinberger, Convolutional networks with dense connectivity, *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019) 1–1doi:10.1109/TPAMI.2019.2918284.
- [50] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications, arXiv preprint arXiv:1704.04861 (2017).
- [51] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241.
- [52] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, *IEEE transactions on pattern analysis and machine intelligence* 39 (12) (2017) 2481–2495.
- [53] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2881–2890.
- [54] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, *IEEE transactions on pattern analysis and machine intelligence* 40 (4) (2017) 834–848.
- [55] L.-C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation, arXiv preprint arXiv:1706.05587 (2017).
- [56] M.-H. Guo, T.-X. Xu, J.-J. Liu, et al., Attention mechanisms in computer vision: A survey, *Computational Visual Media* 8 (3) (2022) 331–368.
- [57] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132–7141.

- [58] J. Park, S. Woo, J.-Y. Lee, I. S. Kweon, Bam: Bottleneck attention module, arXiv preprint arXiv:1807.06514 (2018).
- [59] S. Woo, J. Park, J.-Y. Lee, I. S. Kweon, Cbam: Convolutional block attention module, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 3–19.
- [60] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, Q. Hu, Eca-net: Efficient channel attention for deep convolutional neural networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11534–11542.
- [61] Y. Cao, J. Xu, S. Lin, F. Wei, H. Hu, Gcnet: Non-local networks meet squeeze-excitation networks and beyond, in: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019, pp. 0–0.
- [62] D. Misra, T. Nalamada, A. U. Arasanipalai, Q. Hou, Rotate to attend: Convolutional triplet attention module, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021, pp. 3139–3148.
- [63] J. Huang, M. Tu, W. Yang, W. Kang, Joint attention network for finger vein authentication, *IEEE Transactions on Instrumentation and Measurement* 70 (2021) 1–11.
- [64] T. K. Dutta, D. R. Nayak, Cdancet: Channel split dual attention based cnn for brain tumor classification in mr images, in: 2022 IEEE International Conference on Image Processing (ICIP), IEEE, 2022, pp. 4208–4212.
- [65] A. G. Roy, N. Navab, C. Wachinger, Recalibrating fully convolutional networks with spatial and channel “squeeze and excitation” blocks, *IEEE transactions on medical imaging* 38 (2) (2018) 540–549.
- [66] N. K. Tomar, et al., DDANet: Dual decoder attention network for automatic polyp segmentation, in: International Conference on Pattern Recognition, Springer, 2021, pp. 307–314.
- [67] S. Safarov, T. K. Whangbo, A-denseunet: Adaptive densely connected u-net for polyp segmentation in colonoscopy images with atrous convolution, *Sensors* 21 (4) (2021) 1441.
- [68] P. Sharma, A. Gautam, P. Maji, R. B. Pachori, B. K. Balabantary, Li-segnet: Encoder-decoder mode lightweight segmentation network for colorectal polyps analysis, *IEEE Transactions on Biomedical Engineering* (2022) 1–10doi:10.1109/TBME.2022.3216269.

- [69] A. Duran, G. Dussert, O. Rouvière, T. Jaouen, P.-M. Jodoin, C. Lartizien, Prostattention-net: A deep attention model for prostate cancer segmentation by aggressiveness in mri scans, *Medical Image Analysis* 77 (2022) 102347.
- [70] J. Li, K. Liu, Y. Hu, H. Zhang, A. A. Heidari, H. Chen, W. Zhang, A. D. Algarni, H. Elmannai, Eres-unet++: Liver ct image segmentation based on high-efficiency channel attention and res-unet++, *Computers in Biology and Medicine* 158 (2023) 106501.
- [71] H. Xiao, Q. Liu, L. Li, Mfmanet: Multi-feature multi-attention network for efficient subtype classification on non-small cell lung cancer ct images, *Biomedical Signal Processing and Control* 84 (2023) 104768.
- [72] B. Xu, J. Liu, X. Hou, B. Liu, J. Garibaldi, I. O. Ellis, A. Green, L. Shen, G. Qiu, Attention by selection: A deep selective attention approach to breast cancer classification, *IEEE Transactions on Medical Imaging* 39 (6) (2019) 1930–1941.
- [73] H. Hu, Z. Zhang, Z. Xie, S. Lin, Local relation networks for image recognition, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 3464–3473.
- [74] H. Zhao, J. Jia, V. Koltun, Exploring self-attention for image recognition, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 10076–10085.
- [75] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, in: *International Conference on Learning Representations (ICLR)*, 2021.
- [76] J. Zhang, Y. Xie, Q. Wu, Y. Xia, Medical image classification using synergic deep learning, *Medical Image Analysis* 54 (2019) 10–19.
- [77] M. H. Yap, G. Pons, J. Martí, S. Ganau, M. Sentís, R. Zwiggelaar, A. K. Davison, R. Martí, Automated breast ultrasound lesions detection using convolutional neural networks, *IEEE journal of biomedical and health informatics* 22 (4) (2017) 1218–1226.
- [78] D. R. Nayak, R. Dash, B. Majhi, Automated diagnosis of multi-class brain abnormalities using mri images: a deep convolutional neural network based method, *Pattern Recognition Letters* 138 (2020) 385–391.
- [79] Y. Wang, Z. Feng, L. Song, X. Liu, S. Liu, Multiclassification of endoscopic colonoscopy images based on deep transfer learning, *Computational and Mathematical Methods in Medicine* 2021 (2021).

- [80] A. R. Saikia, K. Bora, L. B. Mahanta, A. K. Das, Comparative assessment of cnn architectures for classification of breast fnac images, *Tissue and Cell* 57 (2019) 8–14.
- [81] L. Zhang, V. Gopalakrishnan, L. Lu, R. M. Summers, J. Moss, J. Yao, Self-learning to detect and segment cysts in lung ct images without manual annotation, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE, 2018, pp. 1100–1103.
- [82] L. Sun, W. Shao, D. Zhang, M. Liu, Anatomical attention guided deep networks for roi segmentation of brain mr images, *IEEE Transactions on Medical Imaging* 39 (6) (2020) 2000–2012. doi:10.1109/TMI.2019.2962792.
- [83] X. Li, H. Chen, X. Qi, Q. Dou, C. W. Fu, P. A. Heng, H-denseunet: Hybrid densely connected unet for liver and tumor segmentation from ct volumes, *IEEE Transactions on Medical Imaging* 37 (12) (2018) 2663–2674. doi:10.1109/TMI.2018.2845918.
- [84] L. Sun, W. Ma, X. Ding, Y. Huang, D. Liang, J. Paisley, A 3d spatially weighted network for segmentation of brain tissue from mri, *IEEE Transactions on Medical Imaging* 39 (4) (2020) 898–909. doi:10.1109/TMI.2019.2937271.
- [85] M. Rajchl, M. C. Lee, O. Oktay, K. Kamnitsas, J. Passerat-Palmbach, W. Bai, M. Damodaram, M. A. Rutherford, J. V. Hajnal, B. Kainz, et al., Deepcut: Object segmentation from bounding box annotations using convolutional neural networks, *IEEE Transactions on Medical Imaging* 36 (2) (2016) 674–683.
- [86] Z. Guo, X. Li, H. Huang, N. Guo, Q. Li, Deep learning-based image segmentation on multimodal medical imaging, *IEEE Transactions on Radiation and Plasma Medical Sciences* 3 (2) (2019) 162–169.
- [87] L. Hou, V. Nguyen, A. B. Kanevsky, D. Samaras, T. M. Kurc, T. Zhao, R. R. Gupta, Y. Gao, W. Chen, D. Foran, et al., Sparse autoencoder for unsupervised nucleus detection and representation in histopathology images, *Pattern recognition* 86 (2019) 188–200.
- [88] N. Wu, J. Phang, J. Park, Y. Shen, Z. Huang, M. Zorin, S. Jastrzebski, T. Fevry, J. Katsnelson, E. Kim, et al., Deep neural networks improve radiologists' performance in breast cancer screening, *IEEE Transactions on Medical Imaging* 39 (4) (2019) 1184–1194.
- [89] A. Saber, M. Sakr, O. M. Abo-Seida, A. Keshk, H. Chen, A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique, *IEEE Access* 9 (2021) 71194–71209.

- [90] Y. Wang, N. Wang, M. Xu, J. Yu, C. Qin, X. Luo, X. Yang, T. Wang, A. Li, D. Ni, Deeply-supervised networks with threshold loss for cancer detection in automated breast ultrasound, *IEEE Transactions on Medical Imaging* 39 (4) (2019) 866–876.
- [91] Y. Luo, Q. Huang, X. Li, Segmentation information with attention integration for classification of breast tumor in ultrasound image, *Pattern Recognition* 124 (2022) 108427.
- [92] Q. Hu, H. M. Whitney, M. L. Giger, A deep learning methodology for improved breast cancer diagnosis using multiparametric mri, *Scientific reports* 10 (1) (2020) 1–11.
- [93] H. Li, D. Chen, W. H. Nailon, M. E. Davies, D. I. Laurenson, Dual convolutional neural networks for breast mass segmentation and diagnosis in mammography, *IEEE Transactions on Medical Imaging* 41 (1) (2021) 3–13.
- [94] C. Li, X. Wang, W. Liu, L. J. Latecki, B. Wang, J. Huang, Weakly supervised mitosis detection in breast histopathology images using concentric loss, *Medical Image Analysis* 53 (2019) 165–178.
- [95] Yadavendra, S. Chand, A comparative study of breast cancer tumor classification by classical machine learning methods and deep learning method, *Machine Vision and Applications* 31 (6) (2020) 46.
- [96] M. Veta, Y. J. Heng, N. Stathonikos, B. E. Bejnordi, F. Beca, T. Wollmann, K. Rohr, M. A. Shah, D. Wang, M. Rousson, et al., Predicting breast tumor proliferation from whole-slide images: the tupac16 challenge, *Medical Image Analysis* 54 (2019) 111–121.
- [97] M. Liu, L. Hu, Y. Tang, C. Wang, Y. He, C. Zeng, K. Lin, Z. He, W. Huo, A deep learning method for breast cancer classification in the pathology images, *IEEE Journal of Biomedical and Health Informatics* 26 (10) (2022) 5025–5032.
- [98] Q. He, Q. Yang, M. Xie, Hctnet: A hybrid cnn-transformer network for breast ultrasound image segmentation, *Computers in Biology and Medicine* 155 (2023) 106629.
- [99] N. Bayramoglu, J. Kannala, J. Heikkilä, Deep learning for magnification independent breast cancer histopathology image classification, in: 2016 23rd International conference on pattern recognition (ICPR), IEEE, 2016, pp. 2440–2445.
- [100] X. Qi, L. Zhang, Y. Chen, Y. Pi, Y. Chen, Q. Lv, Z. Yi, Automated diagnosis of breast ultrasonography images using deep neural networks, *Medical Image Analysis* 52 (2019) 185–198.

- [101] X. Yu, C. Kang, D. S. Guttely, S. Kadry, Y. Chen, Y.-D. Zhang, Resnet-scdas-50 for breast abnormality classification, *IEEE/ACM transactions on computational biology and bioinformatics* 18 (1) (2020) 94–102.
- [102] Q. Dou, H. Chen, L. Yu, J. Qin, P.-A. Heng, Multilevel contextual 3-d cnns for false positive reduction in pulmonary nodule detection, *IEEE Transactions on Biomedical Engineering* 64 (7) (2016) 1558–1567.
- [103] S. E. Gerard, T. J. Patton, G. E. Christensen, J. E. Bayouth, J. M. Reinhardt, Fissurenet: A deep learning approach for pulmonary fissure detection in ct images, *IEEE Transactions on Medical Imaging* 38 (1) (2018) 156–166.
- [104] J. Jiang, Y.-C. Hu, C.-J. Liu, D. Halpenny, M. D. Hellmann, J. O. Deasy, G. Mageras, H. Veeraraghavan, Multiple resolution residually connected feature streams for automatic lung tumor segmentation from ct images, *IEEE Transactions on Medical Imaging* 38 (1) (2018) 134–144.
- [105] D. Moitra, R. K. Mandal, Classification of non-small cell lung cancer using one-dimensional convolutional neural network, *Expert Systems with Applications* (2020) 113564.
- [106] X. Wang, H. Chen, C. Gan, H. Lin, Q. Dou, E. Tsougenis, Q. Huang, M. Cai, P.-A. Heng, Weakly supervised deep learning for whole slide lung cancer image analysis, *IEEE Transactions on Cybernetics* 50 (9) (2019) 3950–3962.
- [107] D. M. Ibrahim, N. M. Elshennawy, A. M. Sarhan, Deep-chest: Multi-classification deep learning model for diagnosing covid-19, pneumonia, and lung cancer chest diseases, *Computers in biology and medicine* 132 (2021) 104348.
- [108] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K.-i. Komatsu, M. Matsui, H. Fujita, Y. Kodera, K. Doi, Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules, *American Journal of Roentgenology* 174 (1) (2000) 71–74.
- [109] C. Tong, B. Liang, Q. Su, M. Yu, J. Hu, A. K. Bashir, Z. Zheng, Pulmonary nodule classification based on heterogeneous features learning, *IEEE Journal on Selected Areas in Communications* 39 (2) (2020) 574–581.
- [110] Y. Xu, L. F. Souza, I. C. Silva, A. G. Marques, F. H. Silva, V. X. Nunes, T. Han, C. Jia, V. H. C. de Albuquerque, P. P. Rebouças Filho, A soft computing automatic based in deep learning with use of fine-tuning for pulmonary segmentation in computed tomography images, *Applied Soft Computing* 112 (2021) 107810.

- [111] J. L. Causey, K. Li, X. Chen, W. Dong, K. Walker, J. A. Qualls, J. Stubblefield, J. H. Moore, Y. Guan, X. Huang, Spatial pyramid pooling with 3d convolution improves lung cancer detection, *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 19 (2) (2022) 1165–1172. doi:10.1109/TCBB.2020.3027744.
- [112] H. Hu, Q. Li, Y. Zhao, Y. Zhang, Parallel deep learning algorithms with hybrid attention mechanism for image segmentation of lung tumors, *IEEE Transactions on Industrial Informatics* 17 (4) (2020) 2880–2889.
- [113] H. Huang, R. Wu, Y. Li, C. Peng, Self-supervised transfer learning based on domain adaptation for benign-malignant lung nodule classification on thoracic ct, *IEEE Journal of Biomedical and Health Informatics* 26 (8) (2022) 3860–3871. doi:10.1109/JBHI.2022.3171851.
- [114] Z. Xu, H. Ren, W. Zhou, Z. Liu, Isanet: Non-small cell lung cancer classification and detection based on cnn and attention mechanism, *Biomedical Signal Processing and Control* 77 (2022) 103773.
- [115] A. Halder, D. Dey, A. K. Sadhu, Lung nodule detection from feature engineering to deep learning in thoracic ct images: a comprehensive review, *Journal of Digital Imaging* (2020) 1–23.
- [116] Y.-R. Van Eycke, C. Balsat, L. Verset, O. Debeir, I. Salmon, C. Decaestecker, Segmentation of glandular epithelium in colorectal tumours to automatically compartmentalise ihc biomarker quantification: A deep learning approach, *Medical Image Analysis* 49 (2018) 35–45.
- [117] M. Akbari, M. Mohrekesh, S. Rafiei, S. R. Soroushmehr, N. Karimi, S. Samavi, K. Najarian, Classification of informative frames in colonoscopy videos using convolutional neural networks with binarized weights, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2018, pp. 65–68.
- [118] J. N. Kather, J. Krisam, P. Charoentong, T. Luedde, E. Herpel, C.-A. Weis, T. Gaiser, A. Marx, N. A. Valous, D. Ferber, et al., Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study, *PLoS medicine* 16 (1) (2019) e1002730.
- [119] M. M. Hasan, N. Islam, M. M. Rahman, Gastrointestinal polyp detection through a fusion of contourlet transform and neural features, *Journal of King Saud University-Computer and Information Sciences* (2020).
- [120] S. Graham, H. Chen, J. Gamper, Q. Dou, P.-A. Heng, D. Snead, Y. W. Tsang, N. Rajpoot, Mild-net: Minimal information loss dilated network for gland

instance segmentation in colon histology images, *Medical Image Analysis* 52 (2019) 199–211.

- [121] M.-J. Tsai, Y.-H. Tao, Deep learning techniques for the classification of colorectal cancer tissue, *Electronics* 10 (14) (2021) 1662.
- [122] N. K. Tomar, D. Jha, S. Ali, H. D. Johansen, D. Johansen, M. A. Riegler, P. Halvorsen, Ddanet: Dual decoder attention network for automatic polyp segmentation, arXiv preprint arXiv:2012.15245 (2020).
- [123] A. B. Hamida, M. Devanne, J. Weber, C. Truntzer, V. Derangère, F. Ghiringhelli, G. Forestier, C. Wemmert, Deep learning for colon cancer histopathological images analysis, *Computers in Biology and Medicine* (2021) 104730.
- [124] X. Huang, L. Zhuo, H. Zhang, Y. Yang, X. Li, J. Zhang, W. Wei, Polyp segmentation network with hybrid channel-spatial attention and pyramid global context guided feature fusion, *Computerized Medical Imaging and Graphics* 98 (2022) 102072.
- [125] G. Liu, Y. Jiang, D. Liu, B. Chang, L. Ru, M. Li, A coarse-to-fine segmentation frame for polyp segmentation via deep and classification features, *Expert Systems with Applications* 214 (2023) 118975.
- [126] C. Wu, C. Long, S. Li, J. Yang, F. Jiang, R. Zhou, Msraformer: Multiscale spatial reverse attention network for polyp segmentation, *Computers in Biology and Medicine* 151 (2022) 106274.
- [127] X. Wang, W. Yang, J. Weinreb, J. Han, Q. Li, X. Kong, Y. Yan, Z. Ke, B. Luo, T. Liu, et al., Searching for prostate cancer by fully automated magnetic resonance imaging classification: deep learning versus non-deep learning, *Scientific reports* 7 (1) (2017) 1–8.
- [128] S. Yoo, I. Gujrathi, M. A. Haider, F. Khalvati, Prostate cancer detection using deep convolutional neural networks, *Scientific reports* 9 (1) (2019) 1–10.
- [129] A. A. Abbasi, L. Hussain, I. A. Awan, I. Abbasi, A. Majid, M. S. A. Nadeem, Q.-A. Chaudhary, Detecting prostate cancer using deep learning convolution neural network with transfer learning approach, *Cognitive Neurodynamics* 14 (4) (2020) 523–533.
- [130] M. Salvi, M. Bosco, L. Molinaro, A. Gambella, M. Papotti, U. R. Acharya, F. Molinari, A hybrid deep learning approach for gland segmentation in prostate histopathological images, *Artificial Intelligence in Medicine* 115 (2021) 102076.
- [131] C. De Vente, P. Vos, M. Hosseinzadeh, J. Pluim, M. Veta, Deep learning regression for prostate cancer detection and grading in bi-parametric mri, *IEEE Transactions on Biomedical Engineering* 68 (2) (2020) 374–383.

- [132] J. Chen, Z. Wan, J. Zhang, W. Li, Y. Chen, Y. Li, Y. Duan, Medical image segmentation and reconstruction of prostate tumor based on 3d alexnet, *Computer Methods and Programs in Biomedicine* 200 (2021) 105878.
- [133] K. He, C. Lian, E. Adeli, J. Huo, Y. Gao, B. Zhang, J. Zhang, D. Shen, Metri-cunet: Synergistic image-and voxel-level learning for precise prostate segmen-tation via online sampling, *Medical Image Analysis* 71 (2021) 102039.
- [134] P. Hambarde, S. Talbar, A. Mahajan, S. Chavan, M. Thakur, N. Sable, Prostate lesion segmentation in mr images using radiomics based deeply supervised u-net, *Biocybernetics and Biomedical Engineering* 40 (4) (2020) 1421–1435.
- [135] J. Silva-Rodríguez, A. Colomer, V. Naranjo, Weglenet: A weakly-supervised convolutional neural network for the semantic segmentation of gleason grades in prostate histology images, *Computerized Medical Imaging and Graphics* 88 (2021) 101846.
- [136] E. Arvaniti, K. S. Fricker, M. Moret, N. Rupp, T. Hermanns, C. Fankhauser, N. Wey, P. J. Wild, J. H. Rueschoff, M. Claassen, Automated gleason grading of prostate cancer tissue microarrays via deep learning, *Scientific reports* 8 (1) (2018) 1–11.
- [137] D. Hoar, P. Q. Lee, A. Guida, S. Patterson, C. V. Bowen, J. Merrimen, C. Wang, R. Rendon, S. D. Beyea, S. E. Clarke, Combined transfer learning and test-time augmentation improves convolutional neural network-based semantic segmen-tation of prostate cancer from multi-parametric mr images, *Computer Methods and Programs in Biomedicine* 210 (2021) 106375.
- [138] Z. Wang, R. Wu, Y. Xu, Y. Liu, R. Chai, H. Ma, A two-stage cnn method for mri image segmentation of prostate with lesion, *Biomedical Signal Processing and Control* 82 (2023) 104610.
- [139] M. Ding, Z. Lin, C. H. Lee, C. H. Tan, W. Huang, A multi-scale channel attention network for prostate segmentation, *IEEE Transactions on Circuits and Systems II: Express Briefs* (2023).
- [140] H. Liu, G. Huo, Q. Li, X. Guan, M.-L. Tseng, Multiscale lightweight 3d segmen-tation algorithm with attention mechanism: Brain tumor image segmentation, *Expert Systems with Applications* 214 (2023) 119166.
- [141] C. C. Taylor, V. O. Millien, J. K. Hou, N. N. Massarweh, Association between inflammatory bowel disease and colorectal cancer stage of disease and survival, *Journal of Surgical Research* 247 (2020) 77–85.

- [142] C. R. Boland, G. E. Idos, C. Durno, F. M. Giardiello, J. C. Anderson, C. A. Burke, J. A. Dominitz, S. Gross, S. Gupta, B. C. Jacobson, et al., Diagnosis and management of cancer risk in the gastrointestinal hamartomatous polyposis syndromes: recommendations from the us multi-society task force on colorectal cancer, *Gastroenterology* 162 (7) (2022) 2063–2085.
- [143] J. Smith-Palmer, C. Takizawa, W. Valentine, Literature review of the burden of prostate cancer in germany, france, the united kingdom and canada, *BMC urology* 19 (1) (2019) 1–16.
- [144] P. C. Albertsen, Prostate cancer screening and treatment: where have we come from and where are we going?, *BJU international* 126 (2) (2020) 218–224.
- [145] Y. Yuan, M. Chao, Y.-C. Lo, Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance, *IEEE transactions on medical imaging* 36 (9) (2017) 1876–1886.
- [146] S. H. Kassani, P. H. Kassani, A comparative study of deep learning architectures on melanoma detection, *Tissue and Cell* 58 (2019) 76–83.
- [147] T. Saba, M. A. Khan, A. Rehman, S. L. Marie-Sainte, Region extraction and classification of skin cancer: A heterogeneous framework of deep cnn features fusion and reduction, *Journal of Medical Systems* 43 (9) (2019) 1–19.
- [148] L. Wei, K. Ding, H. Hu, Automatic skin cancer detection in dermoscopy images based on ensemble lightweight deep learning network, *IEEE Access* 8 (2020) 99633–99647.
- [149] R. Ashraf, S. Afzal, A. U. Rehman, S. Gul, J. Baber, M. Bakhtyar, I. Mehmood, O.-Y. Song, M. Maqsood, Region-of-interest based transfer learning assisted framework for skin cancer detection, *IEEE Access* 8 (2020) 147858–147871.
- [150] M. A. Kadampur, S. Al Riyae, Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images, *Informatics in Medicine Unlocked* 18 (2020) 100282.
- [151] T. Guergueb, M. A. Akhloufi, Melanoma skin cancer detection using recent deep learning models, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2021, pp. 3074–3077.
- [152] M. Shorfuzzaman, An explainable stacked ensemble of deep learning models for improved melanoma skin cancer detection, *Multimedia Systems* (2021) 1–15.

- [153] M. S. Ali, M. S. Miah, J. Haque, M. M. Rahman, M. K. Islam, An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models, *Machine Learning with Applications* 5 (2021) 100036.
- [154] F. Bagheri, M. J. Tarokh, M. Ziaratban, Skin lesion segmentation based on mask rcnn, multi atrous full-cnn, and a geodesic method, *International Journal of Imaging Systems and Technology* 31 (3) (2021) 1609–1624.
- [155] H. C. Reis, V. Turk, K. Khoshelham, S. Kaya, Insinet: a deep convolutional approach to skin cancer detection and segmentation, *Medical & Biological Engineering & Computing* (2022) 1–20.
- [156] H. K. Gajera, D. R. Nayak, M. A. Zaveri, Fusion of local and global feature representation with sparse autoencoder for improved melanoma classification, in: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2022, pp. 5051–5054.
- [157] K. Hu, J. Lu, D. Lee, D. Xiong, Z. Chen, As-net: Attention synergy network for skin lesion segmentation, *Expert Systems with Applications* 201 (2022) 117112.
- [158] M. D. Kumar, G. Sivanarayana, D. Indira, M. P. Raj, Skin cancer segmentation with the aid of multi-class dilated d-net (md2n) framework, *Multimedia Tools and Applications* (2023) 1–24.
- [159] R. Vivanti, A. Szekeskin, N. Lev-Cohain, J. Sosna, L. Joskowicz, Automatic detection of new tumors and tumor burden evaluation in longitudinal liver ct scan studies, *International journal of computer assisted radiology and surgery* 12 (11) (2017) 1945–1957.
- [160] E. Trivizakis, G. C. Manikis, K. Nikiforaki, K. Drevelegas, M. Constantinides, A. Drevelegas, K. Marias, Extending 2-d convolutional neural networks to 3-d for advancing deep learning cancer classification with application to mri liver tumor differentiation, *IEEE journal of Biomedical and Health Informatics* 23 (3) (2018) 923–930.
- [161] A. Das, U. R. Acharya, S. S. Panda, S. Sabut, Deep learning based liver cancer detection using watershed transform and gaussian mixture model techniques, *Cognitive Systems Research* 54 (2019) 165–175.
- [162] F. Özyurt, T. Tuncer, E. Avci, M. Koç, İ. Serhatlioğlu, A novel liver image classification method using perceptual hash-based convolutional neural network, *Arabian Journal for Science and Engineering* 44 (4) (2019) 3173–3182.
- [163] Y. Zhang, X. Pan, C. Li, T. Wu, 3d liver and tumor segmentation with cnns based on region and distance metrics, *Applied Sciences* 10 (11) (2020) 3794.

- [164] J. D. L. Araújo, L. B. da Cruz, J. L. Ferreira, O. P. da Silva Neto, A. C. Silva, A. C. de Paiva, M. Gattass, An automatic method for segmentation of liver lesions in computed tomography images using deep neural networks, *Expert Systems with Applications* 180 (2021) 115064.
- [165] S. Lal, D. Das, K. Alabhy, A. Kanfade, A. Kumar, J. Kini, Nucleisegnet: Robust deep learning architecture for the nuclei segmentation of liver cancer histopathology images, *Computers in Biology and Medicine* 128 (2021) 104075.
- [166] J. Chi, X. Han, C. Wu, H. Wang, P. Ji, X-net: Multi-branch unet-like network for liver and tumor segmentation from 3d abdominal ct scans, *Neurocomputing* 459 (2021) 81–96.
- [167] Q. Gao, M. Almekkawy, Asu-net++: A nested u-net with adaptive feature extractions for liver tumor segmentation, *Computers in Biology and Medicine* (2021) 104688.
- [168] A. Aghamohammadi, R. Ranjbarzadeh, F. Naiemi, M. Mogharrebi, S. Dorost, M. Bendechache, Tpcnn: Two-path convolutional neural network for tumor and liver segmentation in ct images using a novel encoding approach, *Expert Systems with Applications* (2021) 115406.
- [169] R. Ranjbarzadeh, S. B. Saadi, Automated liver and tumor segmentation based on concave and convex points using fuzzy c-means and mean shift clustering, *Measurement* 150 (2020) 107086.
- [170] C. Zhang, J. Lu, Q. Hua, C. Li, P. Wang, Saa-net: U-shaped network with scale-axis-attention for liver tumor segmentation, *Biomedical Signal Processing and Control* 73 (2022) 103460.
- [171] M. Rela, N. R. Suryakari, R. R. Patil, A diagnosis system by u-net and deep neural network enabled with optimal feature selection for liver tumor detection using ct images, *Multimedia Tools and Applications* 82 (3) (2023) 3185–3227.
- [172] S. Deepak, P. Ameer, Brain tumor classification using deep cnn features via transfer learning, *Computers in biology and medicine* 111 (2019) 103345.
- [173] S. A. A. Ismael, A. Mohammed, H. Hefny, An enhanced deep learning approach for brain cancer MRI images classification using residual networks, *Artificial Intelligence in Medicine* 102 (2020) 101779.
- [174] M. A. Naser, M. J. Deen, Brain tumor segmentation and grading of lower-grade glioma using deep learning in mri images, *Computers in biology and medicine* 121 (2020) 103758.

- [175] R. Mehrotra, M. Ansari, R. Agrawal, R. Anand, A transfer learning approach for ai-based classification of brain tumors, *Machine Learning with Applications* 2 (2020) 100003.
- [176] G. S. Tandel, A. Tiwari, O. Kakde, Performance optimisation of deep learning models using majority voting algorithm for brain tumour classification, *Computers in Biology and Medicine* (2021) 104564.
- [177] G. Karayegen, M. F. Aksahin, Brain tumor prediction on mr images with semantic segmentation by using deep learning network and 3d imaging of tumor region, *Biomedical Signal Processing and Control* 66 (2021) 102458.
- [178] Y. Wang, Z. Zhao, S. Hu, F. Chang, Clcu-net: Cross-level connected u-shaped network with selective feature aggregation attention module for brain tumor segmentation, *Computer Methods and Programs in Biomedicine* 207 (2021) 106154.
- [179] M. I. Sharif, J. P. Li, M. A. Khan, M. A. Saleem, Active deep neural network features selection for segmentation and recognition of brain tumors using mri images, *Pattern Recognition Letters* 129 (2020) 181–189.
- [180] X. Zhou, X. Li, K. Hu, Y. Zhang, Z. Chen, X. Gao, Erv-net: An efficient 3d residual neural network for brain tumor segmentation, *Expert Systems with Applications* 170 (2021) 114566.
- [181] Z. Huang, Y. Zhao, Y. Liu, G. Song, Gcaunet: A group cross-channel attention residual unet for slice based brain tumor segmentation, *Biomedical Signal Processing and Control* 70 (2021) 102958.
- [182] Q. Hao, Y. Pei, R. Zhou, B. Sun, J. Sun, S. Li, X. Kang, Fusing multiple deep models for in vivo human brain hyperspectral image classification to identify glioblastoma tumor, *IEEE Transactions on Instrumentation and Measurement* 70 (2021) 1–14. doi:10.1109/TIM.2021.3117634.
- [183] M. A. Ottom, H. A. Rahman, I. D. Dinov, Znet: deep learning approach for 2d mri brain tumor segmentation, *IEEE Journal of Translational Engineering in Health and Medicine* 10 (2022) 1–8.
- [184] F. Yousaf, S. Iqbal, N. Fatima, T. Kousar, M. S. M. Rahim, Multi-class disease detection using deep learning and human brain medical imaging, *Biomedical Signal Processing and Control* 85 (2023) 104875.
- [185] K. Muhammad, S. Khan, J. Del Ser, V. H. C. De Albuquerque, Deep learning for multigrade brain tumor classification in smart healthcare systems: A prospective survey, *IEEE Transactions on Neural Networks and Learning Systems* 32 (2) (2020) 507–522.

- [186] M. I. Razzak, M. Imran, G. Xu, Efficient brain tumor segmentation with multiscale two-pathway-group conventional neural networks, *IEEE journal of Biomedical and Health Informatics* 23 (5) (2018) 1911–1919.
- [187] S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman, et al., The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans, *Medical physics* 38 (2) (2011) 915–931.
- [188] A. P. Reeves, A. M. Biancardi, D. Yankelevitz, S. Fotin, B. M. Keller, A. Jirapatnakul, J. Lee, A public image database to support research in computer aided diagnosis, in: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2009, pp. 3715–3718.
- [189] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, J. S. Cardoso, Inbreast: toward a full-field digital mammographic database, *Academic radiology* 19 (2) (2012) 236–248.
- [190] mini MODAS, The mini-mias database of mammograms, <http://peipa.essex.ac.uk/info/mias.html>. (2022).
- [191] M. Heath, K. Bowyer, D. Kopans, R. Moore, W. P. Kegelmeyer, The digital database for screening mammography, in: Proceedings of the 5th international workshop on digital mammography, Medical Physics Publishing, 2000, pp. 212–218.
- [192] A. Santos, R. Wernersson, L. J. Jensen, Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes, *Nucleic acids research* 43 (D1) (2014) D1140–D1144.
- [193] K. Pogorelov, K. R. Randel, C. Griwodz, S. L. Eskeland, T. de Lange, D. Johansen, C. Spampinato, D.-T. Dang-Nguyen, M. Lux, P. T. Schmidt, et al., Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection, in: Proceedings of the 8th ACM on Multimedia Systems Conference, ACM, 2017, pp. 164–169.
- [194] D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, H. D. Johansen, Kvasir-seg: A segmented polyp dataset, in: International Conference on Multimedia Modeling, Springer, 2020, pp. 451–462.
- [195] K. Pogorelov, K. R. Randel, T. de Lange, S. L. Eskeland, C. Griwodz, D. Johansen, C. Spampinato, M. Taschwer, M. Lux, P. T. Schmidt, et al., Nerthus: A bowel preparation quality video dataset, in: Proceedings of the 8th ACM on Multimedia Systems Conference, ACM, 2017, pp. 170–174.

- [196] N. Tajbakhsh, S. R. Gurudu, J. Liang, Automated polyp detection in colonoscopy videos using shape and context information, *IEEE Transactions on Medical Imaging* 35 (2) (2015) 630–644.
- [197] P. Choyke, B. Turkbey, P. Pinto, M. Merino, B. Wood, Data from prostate-mri, *The Cancer Imaging Archive* (2016).
- [198] P. Tschandl, C. Rosendahl, H. Kittler, The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, *Scientific data* 5 (1) (2018) 1–9.
- [199] IRCAD, Ircad france, <https://www.ircad.fr/research/3dircadb/>. (2022).
- [200] MIDAS, MIDAS - Collection Livers and liver tumors with expert hand segmentations, <https://www.insight-journal.org/midas/collection/view/38>. (2022).
- [201] PAIP, Paip 2019 - grand challenge, <https://paip2019.grand-challenge.org/Dataset/>. (2022).
- [202] A. M. Joshi, D. R. Nayak, D. Das, Y.-D. Zhang, Lims-net: A lightweight multi-scale cnn for covid-19 detection from chest ct scans, *ACM Transactions on Management Information Systems (TMIS)* (2022).
- [203] C. Baur, S. Albarqouni, N. Navab, Melanogans: high resolution skin lesion synthesis with gans, *arXiv preprint arXiv:1804.04338* (2018).
- [204] P. Ganesan, S. Rajaraman, R. Long, B. Ghoraani, S. Antani, Assessment of data augmentation strategies toward performance improvement of abnormality classification in chest radiographs, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2019, pp. 841–844.
- [205] H. Salehinejad, E. Colak, T. Dowdell, J. Barfett, S. Valaee, Synthesizing chest x-ray pathology for training deep convolutional neural networks, *IEEE Transactions on Medical Imaging* 38 (5) (2018) 1197–1206.
- [206] M. Ghaffari, A. Sowmya, R. Oliver, Automated brain tumor segmentation using multimodal brain scans: a survey based on models submitted to the brats 2012–2018 challenges, *IEEE Reviews in Biomedical Engineering* 13 (2019) 156–168.
- [207] T. He, J. N. Fong, L. W. Moore, C. F. Ezeana, D. Victor, M. Divatia, M. Vasquez, R. M. Ghobrial, S. T. Wong, An imageomics and multi-network based deep learning model for risk assessment of liver transplantation for hepatocellular cancer, *Computerized Medical Imaging and Graphics* 89 (2021) 101894.

- [208] M. Misawa, S.-e. Kudo, Y. Mori, K. Hotta, K. Ohtsuka, T. Matsuda, S. Saito, T. Kudo, T. Baba, F. Ishida, et al., Development of a computer-aided detection system for colonoscopy and a publicly accessible large colonoscopy video database (with video), *Gastrointestinal Endoscopy* 93 (4) (2021) 960–967.
- [209] B. Jin, Y. Qu, L. Zhang, Z. Gao, Diagnosing parkinson disease through facial expression recognition: video analysis, *Journal of medical Internet research* 22 (7) (2020) e18697.
- [210] M. Yamada, Y. Saito, H. Imaoka, M. Saiko, S. Yamada, H. Kondo, H. Takamaru, T. Sakamoto, J. Sese, A. Kuchiba, et al., Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy, *Scientific reports* 9 (1) (2019) 1–9.
- [211] A. M. Joshi, D. R. Nayak, Mfl-net: An efficient lightweight multi-scale feature learning cnn for covid-19 diagnosis from ct images, *IEEE Journal of Biomedical and Health Informatics* 26 (11) (2022) 5355–5363.
- [212] P. R. Jeyaraj, E. R. S. Nadar, Medical image annotation and classification employing pyramidal feature specific lightweight deep convolution neural network, *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* (2023) 1–12.
- [213] P. Sharma, D. Das, A. Gautam, B. K. Balabantaray, Lpnet: A lightweight cnn with discrete wavelet pooling strategies for colon polyps classification, *International Journal of Imaging Systems and Technology* 33 (2) (2023) 495–510.
- [214] L. Zhang, Y. Wen, A transformer-based framework for automatic covid19 diagnosis in chest cts, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 513–518.
- [215] C. Matsoukas, J. F. Haslum, M. Söderberg, K. Smith, Is it time to replace cnns with transformers for medical images?, *arXiv preprint arXiv:2108.09038* (2021).
- [216] H. Jiang, Z. Diao, T. Shi, Y. Zhou, F. Wang, W. Hu, X. Zhu, S. Luo, G. Tong, Y.-D. Yao, A review of deep learning-based multiple-lesion recognition from medical images: classification, detection and segmentation, *Computers in Biology and Medicine* (2023) 106726.
- [217] E. Prezioso, S. Izzo, F. Giampaolo, F. Piccialli, G. D. Orabona, R. Cuocolo, V. Abbate, L. Ugga, L. Califano, Predictive medicine for salivary gland tumours identification through deep learning, *IEEE Journal of Biomedical and Health Informatics* 26 (10) (2022) 4869–4879. doi:10.1109/JBHI.2021.3120178.

- [218] Y. Wang, C. Jiang, Y. Wu, T. Lv, H. Sun, Y. Liu, L. Li, X. Pan, Semantic-powered explainable model-free few-shot learning scheme of diagnosing covid-19 on chest x-ray, *IEEE Journal of Biomedical and Health Informatics* 26 (12) (2022) 5870–5882. doi:10.1109/JBHI.2022.3205167.