JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 931 - 947
https://publishoa.com
ISSN: 1309-3452

The architectural development of the convolution neural network and its uses in the medical field

Hind Hatem Ramadhan [1] And Prof. Dr. Qassim Mohammed Hussein [2]

Department of Computer Science, Tikrit University, Iraq
hind.h.ramadhan35523@st.tu.edu.iq
kasimalshamry@tu.edu.iq

ABSTRACT

The emergence of neural networking designs with great performance in computer tasks which are vision-related has sparked interest in radiology artificial intelligence (AI). Radiologists could get advantage from a greater understanding the principles of AI as AI-based software systems become more incorporated into the clinical workflow. Machine learning (ML) is becoming increasingly popular in the domains of medical imaging, radiomics, and medical image analysis. Deep learning is a sort of machine learning that originated in the field of computer vision and has since expanded in popularity across a wide-aspect range of sectors. Deep learning has demonstrated outstanding performance in a variety of fields, including picture classification, object detection, and segmentation. This work gives an overall overview of recent achievements in this area by surveying deep learning architectures and DL approaches used to diagnose disease based on medical images.

Keywords: Deep learning, Convolution neural network, medical imaging.

1. Introduction

CNN (Deep Convolutional Neural Networks) is a kind of neural networking which has won multiple competitions in the fields of computer vision and processing images. Image categorization and segmentation, object identification, processing the video, natural language processing, and the recognition of speech are a few of CNN's interesting application domains. In the field of medical imaging, this technology has sparked an interest. Preliminary research has shown encouraging findings, with comparable performance in various detecting tasks to that of qualified radiologists.[1]

Shallow networks, which have only one input and output layer and no more than one hidden layer between them, are used in traditional machine learning. It is classified as deep learning when a network has more than three levels, including input and output layers. As a result, as the number of hidden layers grows, the network becomes more complex.

The Convolutional Neural Network (CNN) is one of the most important deep learning algorithms, and this study reviews the general concepts of neural network construction and its applications to radiologic image processing.[2]

2. Basic CNN components

Kunihiko Fukushima was the first to present CNN[3]. A typical CNN architecture consists of convolution and pooling layers alternated within one single or more fully linked layers at the end. A completely connected layer can be substituted with a global average pooling layer in some instances. To optimize CNN performance, many controlling units like batch normalization and dropout are implemented in addition to different mapping functions. To create novel architectures and achieve improved performance, the arrangement of CNN components is critical.

2.1. Convolutional layer

The output of linked inputs in the receptive field is determined by such layer, which is the main construction element of a convolutional neural networking[4]. Kernels are combined along the height and width of the information[5, and the dot product between input and filter values is computed. As a result, a 2-D activation map for that filter is generated. Once a given type of feature is seen at a particular geographical place in the input, CNN quickly learns which filters to activate.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 931 - 947
https://publishoa.com
ISSN: 1309-3452

2.2. Pooling layer

The outputs of a neuron in single layer can be combined with the outputs of an only one neuron in the following layer using local or global subsampling layers in CNN.[6]. Its major goal is to reduce the number of parameters and calculations in the model by scaling down the spatial size of the representation[7, 8]. It not only makes the calculations go faster, but it also prevents overfitting. Max pooling is the most prevalent type of pooling layer.

2.3. Entirely interconnected layer

The layers of FC could be normal deep NN layers that seek to generate classification or regression predictions from activations[9]. It functions similarly to a multiplelayer perceptron neural system (MLP). Such a layer collects all of the interconnections to each single activation in the preceding layer, where such activations are into calculation by employing matrix multiplication and a bias counterbalance[10].

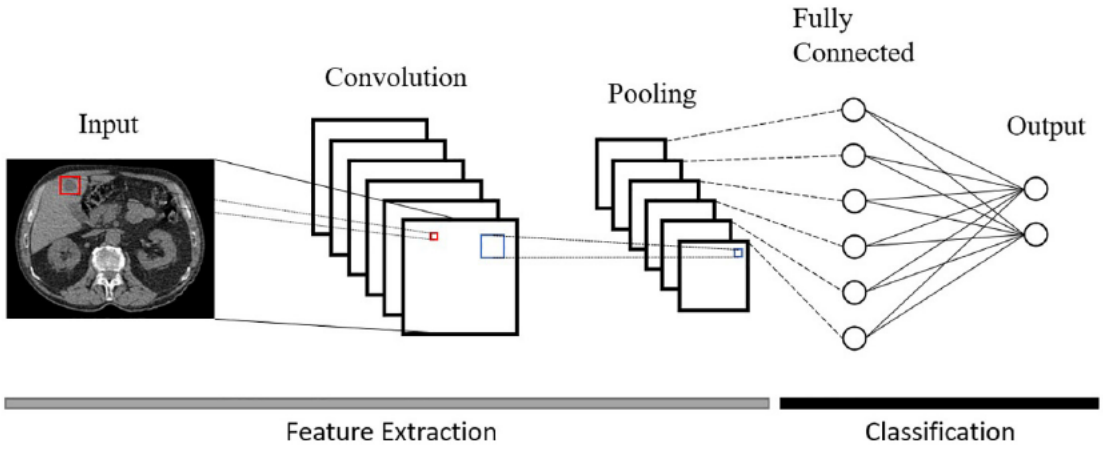

Fig (1): Basic CNN architecture

3. Advantages and limitations of CNN

Some of the advantages of using CNNs in the computer vision environment over other types of neural networks are as follows:

1. The weight sharing aspect, minimizing the amount of trainable networking parameters and so allows the networking to gain generality while avoiding over-fitting, is the main reason to choose CNN.
2. The aspect of extraction layers and the layer of classification must be learned at the same time resulting in a well-organized model output that is dependent on the extracted features.
3. When compared to other neural networks, CNN simplifies the establishment of large-scale networks.[11]

The limitations are:

1. Due to the enormous dimensionality of input data, training has a very high computational cost.
2. the vast amount of training photos required.[12]

4. CNN models

Various advancements in the CNN model have been realized since 1989[2, 13]. These breakthroughs include parameter optimization, structural reformulation, regularization, and other enhancements. However, it appears that the essential driver of CNN performance increase was the reorganization of processing units and the creation of novel blocks. The majority of CNN architectural developments have focused on depth and spatial exploitation. Such kind of exploitation, multiple-path, exploitation of feature-map, depth, width, and attention-based CNNs are the six categories of CNNs depending on the types of architectural changes utilized.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 931 - 947
https://publishoa.com
ISSN: 1309-3452

Figure (2) depicts the taxonomy of CNN architectures graphically. The architectural elements of up-to-date CNN architectures, settings, and performance regarding the datasets of benchmark are summarized in Table (1).

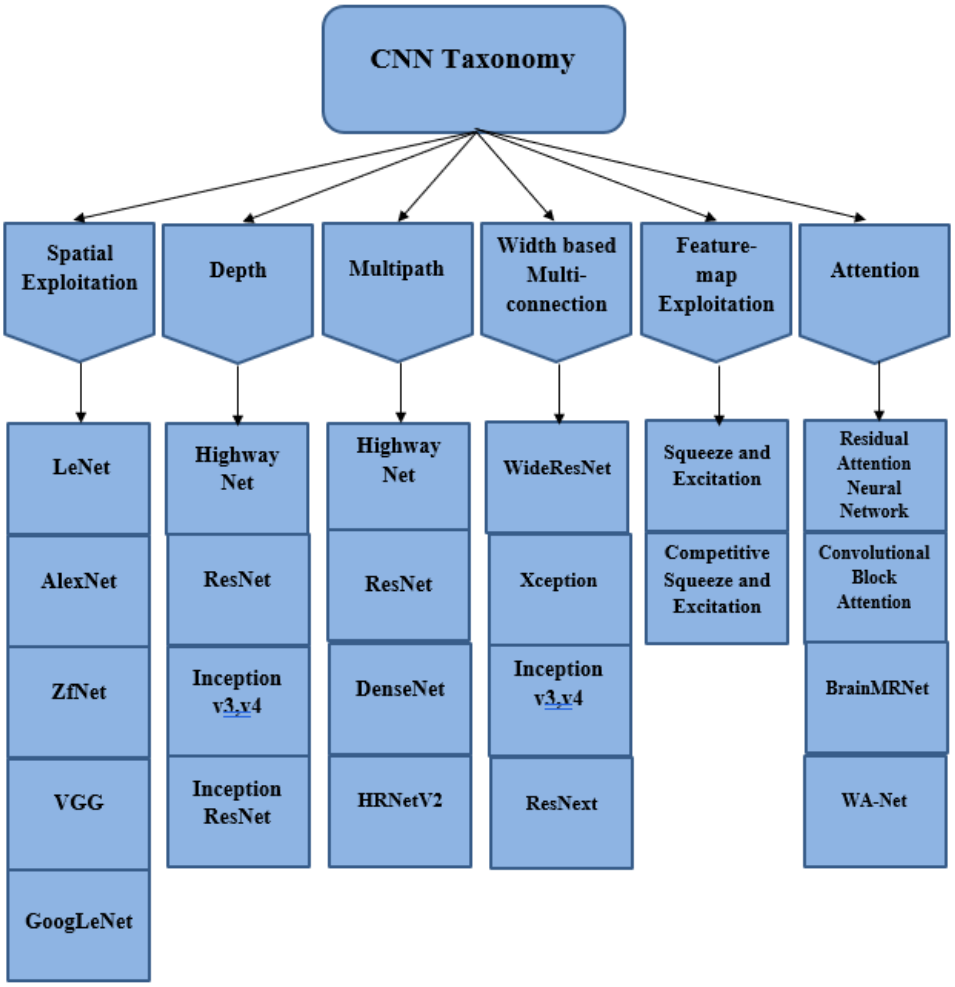

Fig (2): CNN Taxonomy

4.1. Spatial Exploitation based CNN

Weights, biases, processing units (neurons), total of layers, filter size, function of activation, progress, rate of learning, and other parameters and hyper-parameters abound in CNNs.[14, 15]. Because convolutional procedures assess the proximity (locality) of input pixels, different filter sizes may be used to analyze different levels of correlation. Filters of various sizes encompass various levels of granularity; characteristically, small-size filters extract good data while large-size filters retrieve poorly graded data. As a result, in the early 2000s, researchers used spatial filters for increasing performance and investigated the relationship between a spatial filter and network learning. Various studies from the period revealed that by adjusting the filters, CNN may conduct well on all of the coarse and good details.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 931 - 947
https://publishoa.com
ISSN: 1309-3452

4.1.1. LeNet

Yann LeCun dubbed it LeNet-5 after a prior successful iteration[16]. The LeNet architecture was mostly utilized for character recognition jobs like scanning zip codes and numerals, among other things.[17]. The image LeNet input is 32 × 32 × 1 (image by grayscale), that is passed via the convolution layers and afterwards the sub-sampling one. Then there's a pooling layer, which is followed by another set of convolution layers. Finally, there are three layers of FC, with the output layer being one of them. The primary purpose of architecture in post offices was to recognize the patterns of handwritten digit and postal code identification. It makes use of a 5 5filter with a stride of one.[18]

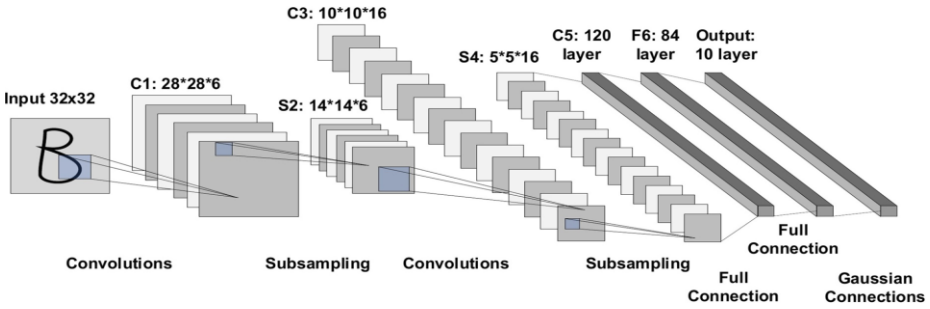

Fig (3): The architecture of LeNet

4.1.2. AlexNet

is among the most important deep CNNs, having won the 2012 ILSVRC competition[19]. The comparable architecture is more complex, with eight layers in total, five of which are convolutional and three of which are fully linked. AlexNet's effective contribution is based on several design and training features.

First, it included the Rectified Linear Unit (ReLU) non-linearity, which assisted in overcoming the vanishing gradient problem and accelerated training. AlexNet also includes a dropout stage, which involves setting a predetermined percentage of layer parameters to zero. To reduce the impact of overfitting, this approach reduces learning parameters and modulates neuron correlation.

Third, when momentum builds, the training process accelerates, and the conditional learning rate decreases (e.g. when learning stagnates). Finally, the number of training data is artificially enhanced by producing versions of the original images that are randomly displaced. As a result, the usage of invariant data representations improves network learning.[20]

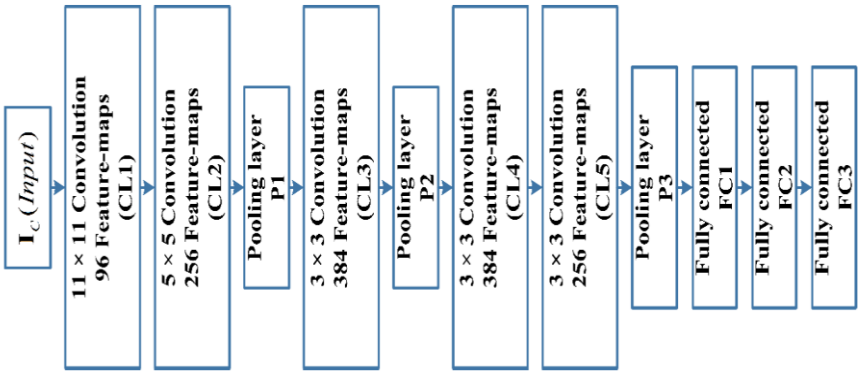

Fig (4): The architecture of AlexNet

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 931 - 947
https://publishoa.com
ISSN: 1309-3452

4.1.3. ZFNet.

Matthew and Rob Fergus proposed it.**[21]**. The ZFNet was the name given to it. It was better than AlexNet because the architecture hyper-parameters were tweaked, especially the size of the intermediate convolutional layers, and the filter size and stride on the first layer were reduced.[17]

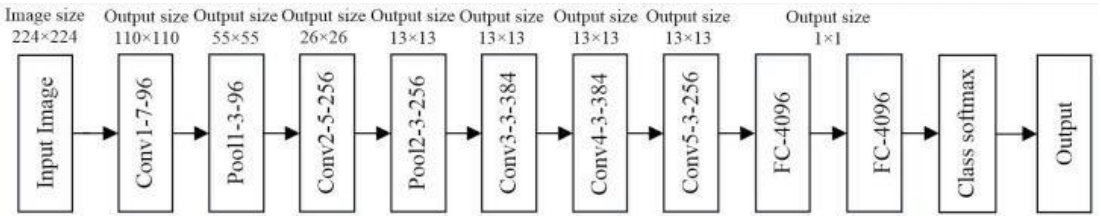

Fig (5): The architecture of ZefNet

4.1.4. VGGNet.

VGG group, Oxford was the runner-up in the ILSVRC 2014 competition.[22] It contains 19 layers in total, which is an improvement over AlexNet. Its key contribution was to demonstrate that network depth, or the number of layers is an important factor in optimal performance. Even though VGGNet achieves fantastic accuracy on the ImageNet dataset, its deployment on even the smallest Graphics Processing Units (GPUs) is difficulty due to massive computing needs in terms of memory and time. Due to the huge breadth of the convolutional layers, it becomes inefficient.[17]

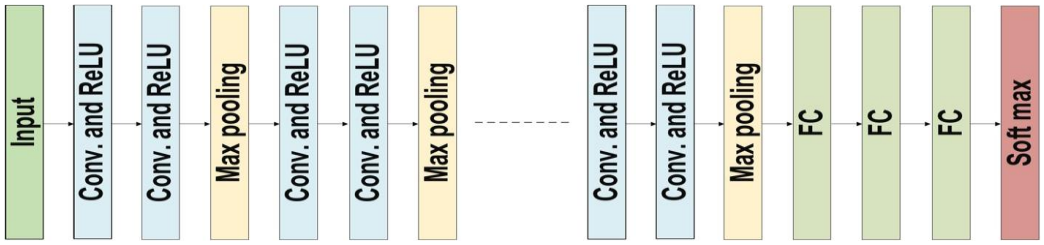

Fig (6): The architecture of VGG

4.1.5. GoogLeNet.

On the ImageNet classification challenge ILSVRC14, GoogleNet achieves the best results.[23]. The increased usage of computer resources in the model is the primary indicator of this model. This is performed using the "inception module," a building element that takes into consideration the model's increased depth and width. This is single kind of first CNN architectures that diverge out of the conventional practice of heaping convolution and pooling layers in sequential order. This network has 22 layers of depth and is substantially quicker than VGG.[18]

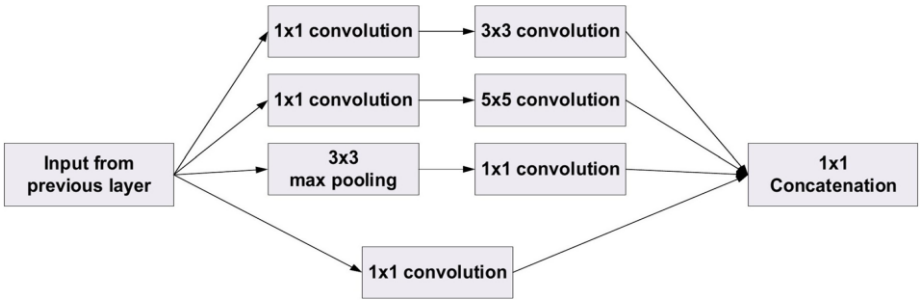

Fig (7): The basic structure of Google Block

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 931 - 947
https://publishoa.com
ISSN: 1309-3452

4.2. Depth-basedCNN

DeepCNN architectures could be centered upon the premise that as the network's depth grows, the more nonlinear mappings and upgraded feature hierarchies it will be able to best mimic the goal function.[24] The network depth has been linked to the effectiveness of supervised training. According to theoretical studies[25]. Deep networking can convey particular kinds of function more effective than superficial structures. Csáji developed a global approximation proposition in 2001, claiming that any function may be estimated using only one unseen layer. This, however, is at the cost of an ever-increasing total of neurons, computationally rendering it unfeasible [26]. In this regard, Bengio proposes that the more deep networking could conserve the communicative power of networking at a lower cost.[27, 28] . is a phrase that can be used to describe a situation. As established by Bengio et al. in 2013[24, 29]. deep networks are computationally more efficient for demanding jobs. The winners of the 2014-ILSVRC competition, Inception and VGG, confirm the concept that network depth is an essential element in regulating learning ability.[22, 23].

4.2.1. ResNet

ResNet [30] is a 152-layer model that combines categorization, localization, and detection into a single spectacular model. Recognize skip interconnections are started. Thus, The inputs of one layer can be copied to the next layer, which overcomes the challenge of training a deep model. This technique is based on the idea that the following layer must learn things fresh and unique from the previous input. ResNet reduces top-1 error by 3.5 percent.

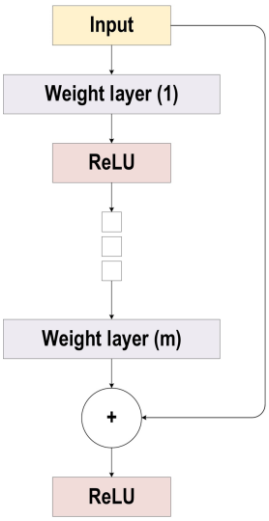

Fig (8): The block diagram for ResNet

4.2.2. Inception V3/V4

The goal of Inception-V3 was to reduce computing costs while maintaining deeper network generalization by using asymmetric filters of small-size (1 × 5 and 1×7) instead of filters of large-size (7 ×7 and 5 ×5); also, they implemented a bottleneck of 1×1 convolution before the filters of large-size. The operation of classic convolution is now extremely similar to cross-channel correlation as a result of these improvements. The input data is divided into three or four spaceswhich are isolated that are tinier than the original input areas using Inception-V3 using the 1 ×1 convolutional technique. Then, using conventional 5 ×5 or 3 ×3 convolutions, all of these correlations are diagramed onto these tiny areas. In Inception, replace The inception block and residual learning power are brought together by the filter concatenation with the residual connection.

4.2.3. Highway Network

Raising networking depth improves performance, especially for complex jobs. The network training, on the other hand, becomes harder. In deeper networks, the presence of multiple layers may lead to tiny gradient values of back-propagation

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 931 - 947
https://publishoa.com
ISSN: 1309-3452

error at lower levels. To address this issue, in 2015 proposed a unique CNN design dubbed Highway Network[31]. This strategy is built on the concept of cross-connectivity. Highway Network enables unrestricted information flow by directing 2 gating units within the layer. The information aggregation was accomplished by combining data from the previous k layers with data from the following k levels, resulting in a regularization effect that simplifies the deeper network's gradient-based training.

4.3 . Multi-Path CNNs

Deep network training is a difficult undertaking, and it has been the focus of contemporary deep network research. Deep CNNs excel at a wide range of tasks. They could, nonetheless, have performance deterioration, gradient disappearance, or explosion concerns as a result of a depth increase rather than overfitting. The fading gradient issue causes a higher test error and a larger training error. The multi-path idea was created to aid in the training of deep neural networks. Multiple shortcut connections or paths can be used to link one layer to another in a logical sequence, skipping certain intermediary levels to allow specialized data to travel between them. [32, 33]. The network is divided into multiple blocks via cross-layer connectivity. These pathways also attempt to tackle the disappearing gradient problem by allowing lower layers to see the gradient. Different forms of shortcut interconnections, like zero-padded, dropout, projection-based, and 1x1 interconnections, are employed for this purpose.

4.3.1. DenseNet

To solve the vanishing gradient problem, DenseNet was developed, and it followed the same route as ResNet and the Highway networking. [34, 35]. Visibly conserving information through preserving individual changes is one of the ResNet problems, even though numerous layers provide very little or no data.

Furthermore, because each layer has its own set of weights, ResNet contains a great number of them. In an improved way of addressing such an issue, DenseNet employed cross-layer connectivity. It used a feed-forward strategy to connect everysingle layer to all other layers in the networking. As a result, the feature maps from each preceding layer had been used as input into the layers that followed.

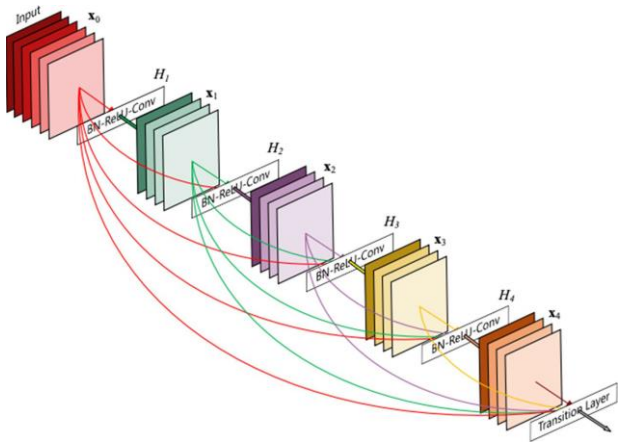

Fig (9): The architecture of DenseNet Network

4.3.2. High-resolution network (HRNet)

Position-sensitive vision processes like semantic segmentation, object recognition, and human posture prediction involve high representations. In today's frameworks, the picture input is encrypted as a depiction with a poor resolution employing a VGGNet or ResNet subnetwork, which is made up of an interconnected series of high-to-low quality convolutions. After that, the low-resolution image is retrieved and transformed to a high-resolution image. A High-Resolution Network, on the other hand, is utilized to retain high-resolution representations throughout the entire process.[36, 37] . Such a network includes two distinct properties. Firstly, the series of convolution from high to low resolution are joined in parallel. Secondly, data is sent back and forth between resolutions on a regular basis. The benefit achieved is a more precise representation in the locative

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 931 - 947
https://publishoa.com
ISSN: 1309-3452

field and a more semantically rich representation in the semantic field. Human posture prediction, object identification, and semantic segmentation are all applications of HRNet., the HRNet is a more strong backbone for computer vision challenges.

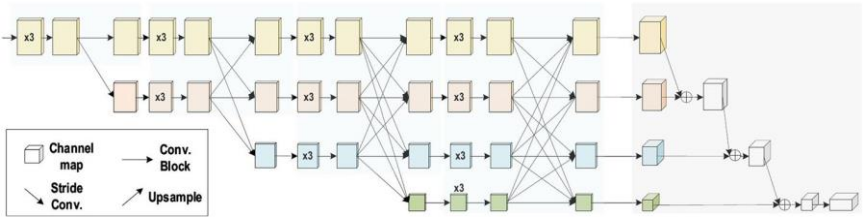

Fig (10): The HRNet general architecture

4.4. Width based Multi-Connection CNNs

Using the potential of depth, as well as the efficacy of multipass interconnections with several passes in network regularization, was a major focus from 2012 to 2015. On the other side, the network's width is critical. Perceptron with many layers compared to perceptron, got the benefit of mapping complicated functions [38]. By using numerous processing units in parallel within a layer. This demonstrates that, like depth, width is critical in the development of learning principles NNs with ReLU activation functions, as recently established, must be broad enough to preserve universal approximation during deepening. Additionally, if the network's maximum width is not greater than the input dimension, an arbitrarily deep network cannot arbitrarily well mimic a class of continuous functions on a compact set. [39]. Although stacking more layers (increasing depth) can assist the NN in learning more diverse feature representations, It does not necessarily improve the learning ability of the NN. One of the most significant disadvantages of deep architectures is the possibility that particular layers or processing units will fail to learn relevant characteristics. Dealing with such issue, researchers have shifted their focus from deep and narrow designs to thin and wide architectures.

4.4.1. WideResNet

Because blocks or transformations of specific feature offer a relatively little amount to learning, the feature reuse problem is the primary drawback associated with deep residual networks[40]. WideResNet proposed a solution to this issue. The depth, according to these scientists, has a supplemental effect, but the rest of units represent deep residual networks' basic learning ability. WideResNet took advantage of the remaining block power by creating the ResNet larger rather than deeper[41]. It increased the width by introducing a new factor, k, that deals with network width.

In other words, comparing to deepening the rest of network, layer broadening is a highly effective method of improving performance. While deep residual networks boost representational capacity, they also have a number of drawbacks, including difficulty with exploding and disappearing gradients, aspect of re-employment (inactivation of numerous feature maps), and the lengthy training process. Adding a dropout among the convolutional layers (rather than within the rest of block) increased learning performance in WideResNet.[42].

4.4.2. Xception

The Xception architecture is a more advanced version of the inception model, consisting completely of depthwise distinct convolutions followed by pointwise convolution[43]. They investigated whether spatial and cross-channel correlations can be sufficiently separated[44]. On ImageNet, this architecture outperforms Inception V3, ResNet, and VGGNet.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 931 - 947
https://publishoa.com
ISSN: 1309-3452

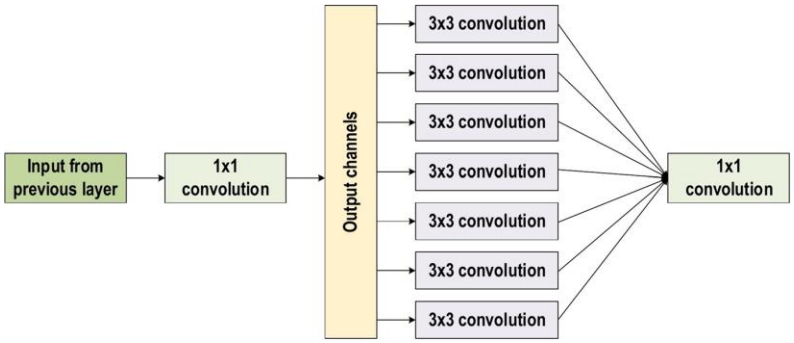

Fig (11): The Xception block architecture's fundamental block diagram.

4.4.3. ResNext

This is a maintainable design that uses the split–transform–merge technique with the ResNet/VGGs method of replicated layers[45]. ResNet/VGGs modify a stack of residual blocks with comparable topologies and rules, resulting in this model. The rules can be moduled as: The blocks are to be split the hyperparameters when the constructed spatial maps have the similar size, and when the spatial maps are assembled by 2, the width of block could be undergone multiplication by 2.

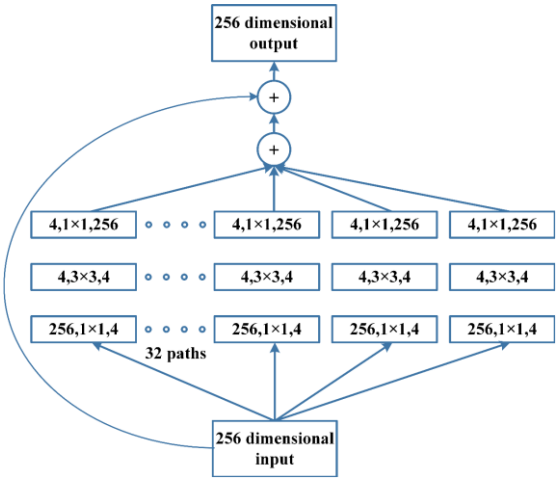

Fig (12): ResNeXt block

4.5. CNNs Feature-Map Exploitation

CNN has become beneficial for MV processes due to its hierarchical learning and feature extraction capabilities[46]. Aspect of selection has a substantial effect on the performance of classification, detection, and segmentation modules. CNN tunes the weights related with a kernel, also recognized as a mask, to choose features dynamically. Furthermore, multiple feature extraction procedures are completed, letting the extraction of a diverse set of characteristics (identified as feature maps or channels in CNN). On the other hand, Some feature maps play a minor or non-existent function in object recognition. Large feature sets may generate a noise effect, leading to network overfitting[47]. This shows that feature-map selection may help with network generalization in addition to network engineering. Many scholars regard feature maps as channels, hence the terms feature maps and channels could be employed interchangeably in this section.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 931 - 947
https://publishoa.com
ISSN: 1309-3452

4.5.1. Squeeze and Excitation

This paradigm is proposed to improve the network's representational power by equipping it for conducting dynamic feature recalibration that is channel-wise [47]. The squeeze and excitation (SE) block were proposed as a novel architectural concept that focuses on the interaction between the channels. SENets laid the framework for the winning ILSVRC 2017 submission of classification, which automatically reduces the top-5 error to 2.251 percent.

4.5.2. Competitive Squeeze and Excitation

In 2018, the CMPESE Network (Competitive Inner Imaging Squeeze and Excitation for Residual Network) was proposed, which used the SEblock principle to improve deep residual network learning[48]. SE-Network re-calibrates the feature maps which are centred on their involvement to class discrimination. The key problem with SE-Net is that ResNet only uses the rest of data to compute the weight of every single feature map. SEblock's impact is reduced, and ResNet data is no longer required. This problem was solved by combining identity mapping with residual feature maps to create feature-map-wise motifs.

4.6. Attention-based CNNs

The discrimination capacity of the NN is determined by several levels of abstraction. For picture recognition and localisation, it's vital to focus on context-relevant features in addition to learning various hierarchies of abstractions. In the human visual system, this phenomenon is known as attention. Man look at the sight in pieces and focus on the portions that are pertinent to the scenario. This approach also deduces numerous interpretations of things at that position, allowing for improved visual structure capture. The interpretability of RNN and LSTM is similar in various aspects. Attention modules are used in RNN and LSTM networks to generate sequential data, and The weighting of fresh samples is determined by their appearance in previous rounds. [49, 50]. In order to improve representation and overcome computational restrictions, some researchers have incorporated the idea of attention into CNN. CNN's capacity to distinguish items despite cluttered backgrounds and complex situations is aided by this concept of attentiveness.

4.6.1. Residual attention neural network

The Residual Attention Network (RAN) has been pr
oposed to accomplish network feature representation[51]. The fundamental goal of paying attention to the CNN is to teach it about the object's conscious qualities. The RAN is a feed-forward CNN since it incorporates amassed residuall blocks as well as the attention unit. The attention module, on the other hand, is split into two sections: the mask branch and the trunk one. Such branches, respectively, utilize top-down and bottom-up learning approaches. Top-down attention feedback and quick feed-forward processing are provided by capturing two distinct methods in the attention model in a single feedforward process[52]. The top-down approach, in particular, generates a dense set of attributes from which to draw conclusions about each piece. Furthermore, In addition to substantial semantic data, the bottom-up feedforward architecture creates low-resolution convolution layer.

4.6.2. Convolutional block attention module

The CBAM (convolutional block attention) module is a new attention-based CNN. This module can be identical to SE-Network in terms of design and functionality. During image classification, SE-Network ignores the spatial localitization of object in the image and only analyzes the contribution of the channels. When it comes to object detection, the spatial placement of the object is crucial. The attention mappings are inferred successively by the convolutional block attention module[53]. To obtain the enhanced feature maps, it employs channel attention first, followed by spatial attention. As in the literature, spatial attention is achieved via 1×1 convolution and pooling functions. The use of a spatial axis in conjunction with feature pooling can be used to create an effective feature descriptor. Furthermore, because CBAM combines the max pooling and the average pooling methods, it is feasible to generate a robust spatial attention map. A set of GAP and max-pooling processes is utilized for modelling the statistics of feature map similarly.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 931 - 947
https://publishoa.com
ISSN: 1309-3452

Table (1): CNN architecture

model	Main contribution	Input size	depth	dataset	Error rate	category	year
LeNet [16]	First CNN architecture	32 × 32 ×1	5	MNIST	0.95	Spatial exploitation	1998
AlexNet [19]	uses Dropout and ReLU, deeper than LeNet	227 × 227 × 3	8	ImageNet	16.4	Spatial exploitation	2012
ZfNet [21]	Visualization idea of intermediate layers	224 × 224 × 3	8	ImageNet	11.7	Spatial exploitation	2014
VGG [22]	small filter size, Increase the depth	224 × 224 × 3	16, 19	ImageNet	7.3	Spatial exploitation	2014
GoogLeNet[23]	Increased the depth used block concept and used concatenation concept	224 × 224 × 3	22	ImageNet	6.7	Spatial exploitation	2015
Highway[31]	Present multipath idea	32 × 32 ×2	19	CIFAR-10	7.76	Depth & Multipath	2015
Inception-V3 [54]	use small filter size, better feature representation	229 × 229 × 3	48	ImageNet	3.5	Depth & Width	2015
Inception V4 [55]	split transform and integration concepts	229 × 229 × 3	70	ImageNet	3.08	Depth & Width	2016
ResNet [30]	Due to symmetry mapping-based skip connections, it is resistant to overfitting	224 × 224 × 3	152	ImageNet	3.57	Depth & Multipath	2016
Inception-ResNet [54]	- Use the idea of residual links	229 × 229 × 3	572	ImageNet	3.52	Depth & Width & Multipath	2016
WideResNet [40]	increased the width and decreased the depth	32 × 32 ×3	28	CIFAR-10 CIFAR-100	3.89 18.85	Width	2016
Xception [43]	A depthwise convolution followed by a pointwise convolution	229 × 229 × 3	71	ImageNet	0.055	width	2017
DenseNet[35]	Layers that are related to each other	224 × 224 × 3	201	CIFAR-10, CIFAR-100	3.46, 17.18,	Multipath	2017

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 931 - 947
https://publishoa.com
ISSN: 1309-3452

	in blocks.			,ImageNet	5.54		
ResNeXt [45]	Grouped convolution		101	ImageNet	4.4	width	2017
Squeeze-and-excitation networks[47]	Modeled interdependencies among channels	229 × 229 × 3	152	ImageNet	2.3	Feature map exploitation	2017
Networks of Competitive squeeze and excitation [48]	Both residual and identity mappings used for re-scaling the channel	32 × 32 ×3	152	CIFAR-10, CIFAR-100	3.58 18.47	Feature map exploitation	2018
Residual attention neural network[51]	Offered the attention mechanism	40 × 40 ×3	452	CIFAR-10, CIFAR-100	3.90 20.4	attention	2017
Convolutional Block Attention [53]	- Exploits both spatial and feature-map information	40 × 40 ×3	101	ImageNet	5.59	Attention	2018
HRNetV2 [36]	High-resolution representations	224 × 224 × 3	-	ImageNet	5.4	Depth & Multipath	2020
3D-CNN-SVM[56]	Three-dimensional convolutional neural networks (3D-CNNs) were applied	3 × 62 × 96× 96	-	ADNI	-	Spatial exploitation	2020
BrainMRNet	hypercolumn technique, attention modules, and residual blocks	-	-	253 MRI	-	Attention	2020
Siamese convolutional neural network (SCNN)[57]	two modified VGG16 parallel layers	224 × 224 × 3	-	OASIS	-	Spatial exploitation	2020

5. Applications of CNN in medical imaging

CNN has achieved great success in the field of diagnosing diseases based on medical images, where their results can be compared with the results of the radiologists themselves. Below are some studies in which different diseases were diagnosed using medical images.

In [58] the researchers were able to diagnose breast cancer by using Shallow-Deep CNN (SD-CNN), A dataset obtained from a tertiary medical facility (Mayo Clinic Arizona) and a publicly available dataset from INbreast are both used and they achieve accuracy 92 %. In 2019 many uses of CNN in medical field, in [59] liver tumor has been diagnosed**,** There are 494 hepatic lesions separated into training (n = 434) and test (n = 60) sets, and a CNN having three convolutional layers, two maximum pooling layers, and two fully connected layers, each with accompanying rectified linear units is utilized. They

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 931 - 947
https://publishoa.com
ISSN: 1309-3452

achieve 92% accuracy, 92% sensitivity, and 98% specificity. In [60] the authors used automated algorithm kernel-based CNN with M-SVM to detect brain tumors by using 40 MRI images**,** effective brain tumor segmentation with low time complexity, low error rate, and high accuracy(84%).In 2020 many papers are used to detect diseases like brain tumor classification using different CNN architecture **[61] [62] [63]** and they achieve high accuracy as shown in table (2), also used to detect Alzheimer's Disease by using 3D-CNN-SVM with accuracy 95.74 [57], and Siamese convolutional neural network (SCNN) with Accuracy 99.05**[56]**. In **[64]**, BrainMRNet was used to detect Brain tumors by using 253 MRI images, this model achieve an accuracy of 96.05. In 2021, a 3D convolutional neural network (3DCNN) was used to distinguish between hepatocellular carcinoma (HCC) and non-HCC lesions based on histological evidence **[65]**. In **[66]**, the author used CNN to detect Left Ventricle Ischemic Scar and achieve an accuracy of 84.7%. Table (2) shows the uses of CNN in the medical field.

Table (2): CNN uses in the medical field

No.	Ref.	year	dataset	aim	method	result.
1	**[58]**	2018	INbreast	Breast Cancer Diagnosis	SD-CNN	Accuracy 92%
2	**[59]**	2019	Train 434 Test 60	liver tumor diagnosis	CNN	accuracy 92% sensitivity 92% specificity 98%
3	**[67]**	2019	1294 647	Classification of Lesions at Breast	CNN	Accuracy 88% Accuracy 83%
4	**[60]**	2019	40 MRI images	Brain Tumor Segmentation	CNN with M-SVM	accuracy 84%
5	**[62]**	2020	Figshare Radiopaedia Harvard	Categorization of Brain Tumor	CNN-SVM	accuracy 95.82% accuracy 99.0% accuracy 98.7%
6	**[61]**	2020	3064 slices	Detection of Brain Tumor	CNN	Accuracy 91.3% Precision 91% Sensitivity 88%
7	**[63]**	2020	21 patients	Brain tumor segmentation	3D Mask R-CNN	Precision 90% Recall 91%
8	**[68]**	2020	50 schwannoma 34 meningioma.	Differentiating Between Spinal Schwannoma and Meningioma	CNN	accuracy 87 %
9	**[69]**	2020	154 3D LGE-MRIs	Segmenting the Left Atrium	double sequentially used CNNs	Accuracy 93.2
10	[57]	2020	ADNI	Detection of Alzheimer's Disease	3D-CNN-SVM	Accuracy 95.74
11	**[56]**	2020	OASIS	Classification of Alzheimer's Disease	Siamese convolutional neural network (SCNN)	Accuracy 99.05
12	**[64]**	2020	253 MRI	Brain tumor	BrainMRNet	Accuracy 96.05
13	**[65]**	2021	93 HCC	distinguish hepatocellular	3D convolutional	accuracy of 87.3%

			57 non-HCC	carcinoma (HCC) and non-HCC lesions based on histological evidence	neural network (CNN)	
14	**[66]**	2021	200 train 25 test	Left Ventricle Ischemic Scar Detection	CNN	Accuracy 84.7%

6. Conclusion

We reviewed the past literature on convolution neural network (CNN) applications in the process of imaging which is medical in this overview paper. We mainly concentrated on their structure and presented more details about their architectures and structures due to the importance and increased issue spaces of CNNs in recent years. The merits and limitations of CNN were identified, as well as the network's evolution and developments, and how CNN is employed in the medical field.

References

1. Mutasa, S., S. Sun, and R. Ha, *Understanding artificial intelligence based radiology studies: CNN architecture.* Clinical Imaging, 2021. **80**: p. 72-76.
2. Khan, A., et al., *A survey of the recent architectures of deep convolutional neural networks.* Artificial intelligence review, 2020. **53**(8): p. 5455-5516.
3. Fukushima, K., *A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position.* Biol. Cybern., 1980. **36**: p. 193-202.
4. Goodfellow, I., Y. Bengio, and A. Courville, *Deep learning*. 2016: MIT press.
5. Bouvrie, J., *Notes on convolutional neural networks.* 2006.
6. Lee, C.-Y., P.W. Gallagher, and Z. Tu. *Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree*. in *Artificial intelligence and statistics*. 2016. PMLR.
7. Wang, T., et al. *End-to-end text recognition with convolutional neural networks*. in *Proceedings of the 21st international conference on pattern recognition (ICPR2012)*. 2012. IEEE.
8. He, K., et al., *Spatial pyramid pooling in deep convolutional networks for visual recognition.* IEEE transactions on pattern analysis and machine intelligence, 2015. **37**(9): p. 1904-1916.
9. Lin, M., Q. Chen, and S. Yan, *Network in network.* arXiv preprint arXiv:1312.4400, 2013.
10. Rawat, W. and Z. Wang, *Deep convolutional neural networks for image classification: A comprehensive review.* Neural computation, 2017. **29**(9): p. 2352-2449.
11. Alzubaidi, L., et al., *Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions.* Journal of big Data, 2021. **8**(1): p. 1-74.
12. Suzuki, K., *Overview of deep learning in medical imaging.* Radiological physics and technology, 2017. **10**(3): p. 257-273.
13. Shrestha, A. and A. Mahmood, *Review of deep learning algorithms and architectures.* IEEE access, 2019. **7**: p. 53040-53065.
14. Kafi, M., M. Maleki, and N. Davoodian, *Functional histology of the ovarian follicles as determined by follicular fluid concentrations of steroids and IGF-1 in Camelus dromedarius.* Research in veterinary science, 2015. **99**: p. 37-40.
15. Lu, L., et al., *Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning deep convolutional neural networks for computer-aided detection: CNN*

architectures, dataset characteristics and transfer. IEEE Transactions on Medical Imaging, 2016. **35**(1285-1298): p. 1602.03409.

16. LeCun, Y., et al., *Gradient-based learning applied to document recognition.* Proceedings of the IEEE, 1998. **86**(11): p. 2278-2324.
17. Coşkun, M., et al., *An overview of popular deep learning methods.* European Journal of Technique (EJT), 2017. **7**(2): p. 165-176.
18. Dhillon, A. and G.K. Verma, *Convolutional neural network: a review of models, methodologies and applications to object detection.* Progress in Artificial Intelligence, 2020. **9**(2): p. 85-112.
19. Krizhevsky, A., I. Sutskever, and G.E. Hinton, *Imagenet classification with deep convolutional neural networks.* Advances in neural information processing systems, 2012. **25**: p. 1097-1105.
20. Jaafra, Y., et al., *Reinforcement learning for neural architecture search: A review.* Image and Vision Computing, 2019. **89**: p. 57-66.
21. Zeiler, M.D. and R. Fergus. *Visualizing and understanding convolutional networks*. in *European conference on computer vision*. 2014. Springer.
22. Simonyan, K. and A. Zisserman, *Very deep convolutional networks for large-scale image recognition.* arXiv preprint arXiv:1409.1556, 2014.
23. Szegedy, C., et al. *Going deeper with convolutions*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
24. Bengio, Y. *Deep learning of representations: Looking forward*. in *International conference on statistical language and speech processing*. 2013. Springer.
25. Montúfar, G., et al., *On the number of linear regions of deep neural networks.* arXiv preprint arXiv:1402.1869, 2014.
26. Csáji, B.C., *Approximation with artificial neural networks (MSc thesis).* Faculty of sciences ötvös Loránd, University Hungary, 2001.
27. Delalleau, O. and Y. Bengio, *Shallow vs. deep sum-product networks.* Advances in neural information processing systems, 2011. **24**: p. 666-674.
28. Hansen, E., et al., *Strong expression of foreign genes following direct injection into fish muscle.* FEBS letters, 1991. **290**(1-2): p. 73-76.
29. Nguyen, Q., M.C. Mukkamala, and M. Hein. *Neural networks should be wide enough to learn disconnected decision regions*. in *International Conference on Machine Learning*. 2018. PMLR.
30. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
31. Srivastava, R.K., K. Greff, and J. Schmidhuber, *Highway networks.* arXiv preprint arXiv:1505.00387, 2015.
32. Mao, X., C. Shen, and Y.-B. Yang, *Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections.* Advances in neural information processing systems, 2016. **29**.
33. Tong, T., et al. *Image super-resolution using dense skip connections*. in *Proceedings of the IEEE international conference on computer vision*. 2017.
34. Wu, S., S. Zhong, and Y. Liu, *Deep residual learning for image steganalysis.* Multimedia tools and applications, 2018. **77**(9): p. 10437-10453.
35. Huang, G., et al. *Densely connected convolutional networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
36. Wang, J., et al., *Deep high-resolution representation learning for visual recognition.* IEEE transactions on pattern analysis and machine intelligence, 2020. **43**(10): p. 3349-3364.
37. Cheng, B., et al. *Higherhrnet: Scale-aware representation learning for bottom-up human pose estimation*. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.
38. Kawaguchi, K., J. Huang, and L.P. Kaelbling, *Effect of depth and width on local minima in deep learning.* Neural computation, 2019. **31**(7): p. 1462-1498.

39. Hanin, B. and M. Sellke, *Approximating continuous functions by relu nets of minimal width.* arXiv preprint arXiv:1710.11278, 2017.
40. Zagoruyko, S. and N. Komodakis, *Wide residual networks.* arXiv preprint arXiv:1605.07146, 2016.
41. Huang, G., et al. *Deep networks with stochastic depth*. in *European conference on computer vision*. 2016. Springer.
42. Huynh, H.T. and H. Nguyen, *Joint age estimation and gender classification of Asian faces using wide ResNet.* SN Computer Science, 2020. **1**(5): p. 1-9.
43. Chollet, F. *Xception: Deep learning with depthwise separable convolutions*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
44. Lo, W.W., X. Yang, and Y. Wang. *An xception convolutional neural network for malware classification with transfer learning*. in *2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*. 2019. IEEE.
45. Xie, S., et al. *Aggregated residual transformations for deep neural networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
46. LeCun, Y., K. Kavukcuoglu, and C. Farabet. *Convolutional networks and applications in vision*. in *Proceedings of 2010 IEEE international symposium on circuits and systems*. 2010. IEEE.
47. Hu, J., L. Shen, and G. Sun. *Squeeze-and-excitation networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
48. Hu, Y., et al., *Competitive inner-imaging squeeze and excitation for residual network.* arXiv preprint arXiv:1807.08920, 2018.
49. Sundermeyer, M., R. Schlüter, and H. Ney. *LSTM neural networks for language modeling*. in *Thirteenth annual conference of the international speech communication association*. 2012.
50. Mikolov, T., et al. *Recurrent neural network based language model*. in *Interspeech*. 2010. Makuhari.
51. Wang, F., et al. *Residual attention network for image classification*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
52. Salakhutdinov, R. and H. Larochelle. *Efficient learning of deep Boltzmann machines*. in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. 2010. JMLR Workshop and Conference Proceedings.
53. Woo, S., et al. *Cbam: Convolutional block attention module*. in *Proceedings of the European conference on computer vision (ECCV)*. 2018.
54. Szegedy, C., et al. *Rethinking the inception architecture for computer vision*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
55. Szegedy, C., et al. *Inception-v4, inception-resnet and the impact of residual connections on learning*. in *Thirty-first AAAI conference on artificial intelligence*. 2017.
56. Feng, W., et al., *Automated MRI-based deep learning model for detection of Alzheimer's disease process.* International Journal of Neural Systems, 2020. **30**(06): p. 2050032.
57. Mehmood, A., et al., *A deep Siamese convolution neural network for multi-class classification of Alzheimer disease.* Brain sciences, 2020. **10**(2): p. 84.
58. Gao, F., et al., *SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis.* Computerized Medical Imaging and Graphics, 2018. **70**: p. 53-62.
59. Hamm, C.A., et al., *Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI.* European radiology, 2019. **29**(7): p. 3338-3347.
60. Thillaikkarasi, R. and S. Saravanan, *An enhancement of deep learning algorithm for brain tumor segmentation using kernel based CNN with M-SVM.* Journal of medical systems, 2019. **43**(4): p. 1-7.
61. Sarkar, S., et al., *A CNN based approach for the detection of brain tumor using MRI scans.* Test Engineering and Management, 2020. **83**: p. 16580-16586.

62. Deepak, S. and P. Ameer, *Automated categorization of brain tumor from mri using cnn features and svm.* Journal of Ambient Intelligence and Humanized Computing, 2021. **12**(8): p. 8357-8369.
63. Jeong, J., et al., *Brain tumor segmentation using 3D Mask R-CNN for dynamic susceptibility contrast enhanced perfusion imaging.* Physics in Medicine & Biology, 2020. **65**(18): p. 185009.
64. Toğaçar, M., B. Ergen, and Z. Cömert, *Tumor type detection in brain MR images of the deep model developed using hypercolumn technique, attention modules, and residual blocks.* Medical & Biological Engineering & Computing, 2021. **59**(1): p. 57-70.
65. Oestmann, P.M., et al., *Deep learning–assisted differentiation of pathologically proven atypical and typical hepatocellular carcinoma (HCC) versus non-HCC on contrast-enhanced MRI of the liver.* European Radiology, 2021. **31**(7): p. 4981-4990.
66. O'Brien, H., et al., *Automated left ventricle ischemic scar detection in CT using deep neural networks.* Frontiers in cardiovascular medicine, 2021. **8**.
67. Truhn, D., et al., *Radiomic versus convolutional neural networks analysis for classification of contrast-enhancing lesions at multiparametric breast MRI.* Radiology, 2019. **290**(2): p. 290-297.
68. Maki, S., et al., *A deep convolutional neural network with performance comparable to radiologists for differentiating between spinal schwannoma and meningioma.* Spine, 2020. **45**(10): p. 694-700.
69. Xiong, Z., et al., *A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging.* Medical Image Analysis, 2021. **67**: p. 101832.