IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11159

Introduction to CNN (Convolution Neural Network) for medical e-Diagnosis

Vicky[1], Mayank Parashar[2]

Student, Computer Science & Engineering, HMR Institute of Technology & Management, Delhi, India[1]

Associate professor, Computer Science & Engineering, HMR Institute of Technology & Management, Delhi, India [2]

Abstract: CNN classification is a big name in the field of image classification, segmentation of images and Natural Language Processing (NLP). The processing of medical images can be done by CNN to diagnose diseases. We will study the working of CNN and every individual component's working. Along with that, we will analyze other researchers' work in this field and review their work in this paper. We will work on the solution to the real-life problem of the lack of healthcare professionals and the high cost of healthcare facilities. We will discuss how a multilayer perceptron approach can solve this real-life problem with computational ability.

Keywords*:* CNN, Convolution Neural Network, medical e-Diagnosis, computer vision, deep learning

I. INTRODUCTION

As we are moving ahead in the 21st century the demand for automation is high. Automation can power all the repetitive tasks which are tedious and vulnerable to threats from human errors. In some of the most developed countries, a very famous and well-known crisis is ongoing for a long time i.e. "lack of working professionals". The countries like the USA has 2 million worker shortages in the manufacturing sector (research from Deloitte and the Manufacturing Institute).

"Our research estimates that the cumulative skills gap — or the positions that likely won't be filled due to a lack of skilled workers — will grow to 2 million between 2015 and 2025", said Craig Giffi, Deloitte's vice-chairman.

Hence we need to employ more and more machines to work in factories, IT offices, shops, supermarkets, and hospitals etc. We need to automate the processes in every sector to reduce the load on humans. If we, as an example, consider a pizza delivery boy, the delivery boy is transporting pizza from one place to another this type of work might sound boring, simple and repetitive. On the other hand, some of the essential workplaces like hospitals which are operating for saving some of the key resources like human beings are lacking workers, so why not to give a chance to machines. Why we can't hire or develop a machine to do it? That same pizza delivery boy can be used somewhere else where we need human intelligence but in pizza delivery; we hardly need any intelligence to do it.

Automation of this type of task can be very hectic and difficult to operate by a machine. As in the 21st century, artificial intelligence is conquering some of the major problems. Especially, deep learning's multilevel perceptron is making a clear win at discovering complex structures in multi-high-dimensional data. In deep learning the algorithm works on multiple levels, more levels mean deeper. So, how deep is deep learning? There is no agreement among the authors or experts. According to Dr Adrian Rosebrock, any deep neural network with more than two hidden layers can be called a deep neural network.

Deep learning is a field that comes under machine learning (ML) and ML further comes under artificial intelligence. Artificial intelligence provides us with several algorithms and search techniques to tackle some of the very common problems. We are more focused on repetitive tasks and boring tasks which may cost very little or no effort from the human side but it can be aggressively hard and complex for artificial intelligence machines [1].

"Deep learning methods are representation-learning methods with multiple levels of representation, obtained by composing simple but nonlinear modules that each transforms the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level. The key aspect of deep learning is that these layers are not designed by human engineers: they are learned from data using a general-purpose learning procedure" – Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, Nature 2015.

Deep learning is also called the representation learning method. This method is a set of methods that make a machine take raw data as input and automatically come across the representations required for classification purposes. It is a technique of multiple levels of representation, which is obtained by composing simple and non-linear modules that each module reconstructs the representation at each level. With the help of these reconstructions, complex problems can also be solved. It is helpful to rectify the feature of input data that are important for classification and put down the irrelevant variations. It is the key feature of deep learning that this learning method learns by using general-purpose learning procedures from data [2].

Deep learning can be categorized into three different categories depending on how the techniques and architectures are planned to use. First Category is Deep Networks for unsupervised learning: when no information about goal class labels is available. It is focused to arrest the high order correlation of the observed data to analyze the pattern among them. The second category is deep learning for supervised learning: in this, the model provided the direct discriminative power for pattern classification. The data we fed are labelled. The third category is Hybrid Deep Network: in this, the goal is to assist discrimination by the results of the unsupervised deep networks. This can be done by better optimization or rectification of deep networks in the supervised learning or it can be done by using the discriminative measure for supervised learning for estimating the parameters in any unsupervised learning networks [3].

In this article, we will be focusing on medical e-diagnosis, how deep learning neural network is responsible for advanced growth in this field. We will be studying the convolution neural network (CNN) in detail. CNN is designed to work on the information which is in the form of multiple numeric arrays numerical arrays just like images, videos, signal & sequences etc [2].

I. WHAT IS CNN?

Moving ahead in the article now we are on the stage to discuss the algorithm we will be using to do the medical e-diagnosis. Convolution Neural Network (CNN, ConvNets) is another form of Artificial Neural Network (ANN). You can read about the ANN in detail in []. This algorithm can be used to solve one of the most complex problems of pattern recognition in image and video. Because CNN is traditionally a form of the ANN so it can optimize itself by learning. While ANN struggles with the high computation power requirement on the other hand the CNN empower us to generate image-based features in an architecture hence forming the network more worthy to the image-based problems in low computation power requirement [4]. Also, ANN struggles with the localisation of the object in the image but CNN doesn’t.

A. *CNN based architecture*

Note: A tensor is a generalised form of a matrix and it is a type of data structure majorly used in Artificial Intelligence. It is implemented as an n-dimensional array. A single dimensional matrix is a vector, two dimensional is a matrix and it can go to higher dimensions as well which we would need to represent the image.

Now let’s discuss the architecture of CNN, the algorithm takes a tensor as an input and it can be any order the CNN can handle it. But usually, the CNN takes the 3rd order tensor as input. Then the input tensor goes through multiple processing each processing denote a layer, like ReLu layer, pooling layer or a loss layer etc.

$$x^1 \rightarrow \mathbf{w^1} \rightarrow x^2 \rightarrow \cdots \rightarrow x^{L-1} \rightarrow \mathbf{w^{L-1}} \rightarrow x^{L} \rightarrow \mathbf{w^L} \rightarrow z$$

It is the basic structure of a CNN showing that the CNN runs layer by layer. **x** is an input(could be a 3rd order tensor). And L is the number of layers and the $\mathbf{w^L}$ shows the parameters involved in the **L**th layer’s processing. The output of the very first layer is $\mathbf{x^2}$. At the same time, it is the output for the second layer as well. The process goes on further until the last layer doesn’t process the input for it i.e. $\mathbf{x^{L-1}}$.

The last layer plays an important role here, as **t** is the target or ground truth for the input, it can compare the difference between the $\mathbf{x^L}$ and **t**, it is also called the loss layer [5].

The variable t represents the ground truth or target which is transformed to a C dimensional vector t, now C and t are the probability mass function and cross-entropy loss can compute the distance among them. The probability mass function is the function that gives us the probability of a random variable when that random variable is equal to a value. Let **Y** be a discrete random variable with range $\mathbf{S_y = \{y_1, y_2, y_3, y_4, \ldots\}}$. The Probability mass function would be: $\mathbf{P_y(y_k) = P(Y{=}y_k)}$, for $\mathbf{k = 1,2,3\ldots}$

There are mainly three layers in CNN. These are the convolution layer, pooling layer and fully connected layer.

B. *Convolution Layer*

The convolution layer is the first layer of the whole Network also it is the key layer. It plays an important role in feature map generation. The parameters of the convolution layer consist of filters which are also called the Kernels or K learnable filters [1], [4]. Kernels are the small array of numbers that can be implemented by the tensor [6]. It depends on how many K filters we are applying then it will generate the same number of feature maps as the applied filters [7].

DOI: 10.17148/IJARCCE.2022.11159

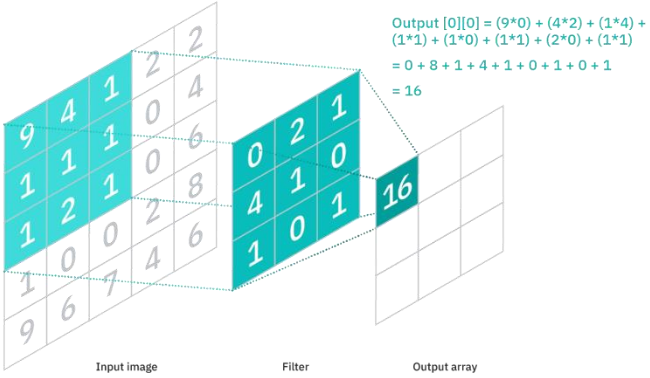

Output [0][0] = (9*0) + (4*2) + (1*4) + (1*1) + (1*0) + (1*1) + (2*0) + (1*1)
= 0 + 8 + 1 + 4 + 1 + 0 + 1 + 0 + 1
= 16
16
Input image
Filter
Output array

Figure 1: Use of filter in CNN [Source: IBM Blog]

Feature Map: When the kernel moves through the input tensor, top to bottom and left to right, and stopping at each coordinate, calculate the element-wise multiplication product and then its sum as the output. It is stored in a feature map [1].

Process of Convolution Layer: In the convolution layer we convolve the filters over the whole image and the feature map is generated as a product. The number of feature maps depends on the number of filters applied. This filter can be small in size but it moves through the entire depth of the input image. Then this feature map is treated by the activation layer. Hence generate the activation map. We will study the activation layer further [4], [7].

The convolution neural network is supposed to have very little complexity as compared to ANN. The convolution layer is itself responsible to reduce or optimizing the size of the output. Some hyperparameters are responsible to manage the size of the output volume. These are **Depth**, **Stride**, and **zero-padding** [1], [4], [6]. A short discussion is below:

Note: The hyperparameters represent a variable that is supposed to be set before the model training starts [6].

C. *Depth*

A depth is a hyperparameter for we can assign a value to optimize the number of neurons that belongs to the same region of input. In the case of ANN, every single neuron in the hidden layer is connected to every single neuron of its predecessor. Lowering the value of the depth hyperparameter can help to reduce the neurons with a significant number. But when we are working with the depth hyperparameter then we have to keep one thing in mind reducing this parameter will also affect the quality of the model negatively [1], [4].

D. *Stride*

In the convolution operation, a kernel or filter or a small matrix is used to slide over the input tensor or input image. That kernel can stop at every coordinate in the larger matrix hence we can have the overlapping receptive region which can cost us more computation power and the quality of model training will be high. But we can manipulate the overlapping region with the help of **stride**, in order to reduce the complexity and decrease the parameters. Stride controls the numbers of coordinates jumped by the kernel when applying operation. Lower stride value will generate more overlapping hence the bigger output volume and higher output value will decrease the overlapping regions and the size of the output volume as well. For example, if we use the stride = 1, then the kernel will convolve through one pixel at a time. If the stride = 2, then the kernel will skip two pixels at a time. It can simply decrease the computational power required. In further this article we will study the pooling layer which is also responsible for optimizing the input size [1], [4], [6], [8], [9].

E. *Zero-padding*

After applying the convolution on the input tensor the output generated after the operation suffers from the reduced height and weight. If the dimension of the output will reduce after every convolutional operation then we can't go into more deeply. Training the model more deeply will not be possible. So here comes the zero-padding to save us. With the help of zero padding, we can add the padding of zeros around the input tensor. For example, if we have a 5x5 input matrix then by adding P = 1 padding we can make it to a 7x7 matrix which will eventually generate the output of the 5x5 matrix same as the original input matrix before. So we can overcome the problem of reduced output matrix hence we would be able to train the networks deeper [4], [6], [8], [9].

These three hyperparameters we just studied can be very useful when it comes to medical e-diagnosis.
Note: A convolution layer’s work can be understood step by step visually by proper figures in a very intuitive manner in the [7].

F. Activation Layer/Non-Linearity Layer

In some cases, it is assumed that the activation immediately occurs after the convolution layer. In the activation layer the output of the linear convolution layer process through an activation function majorly the ReLU (Rectified Linear Unit) is used as an activation function. This layer is used to manipulate the generated output from the previous layer. Earlier the sigmoid and hyperbolic tangents were used as an activation function but ReLU is more famous now due to its simplicity.

$$ReLU(y) = \max(0, y)$$

$$\frac{d}{dy} ReLU(y) = \{1 \; if \; y > 0; 0 \; otherwise\}$$

Here is an example to apply the ReLU function(figure 1):

-156	226	89
254	-234	-125
-265	179	112

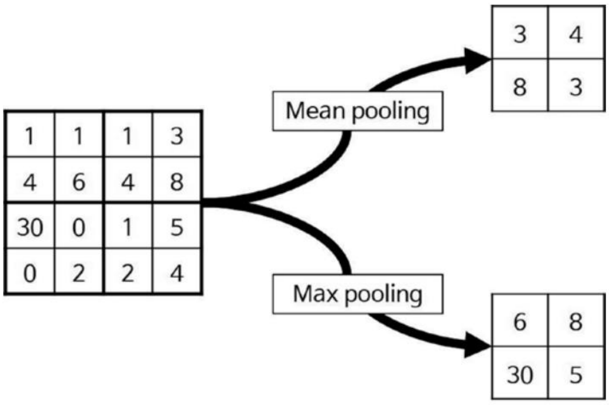
→After Applying ReLU→

0	226	89
254	0	0
0	179	112

Figure 2: Working of ReLU

G. *Pooling Layer*

A pooling layer is usually applied between the successive convolution layers, this layer decreases the size of the input volume and also decrease the number of learnable parameters. The pooling layer decreases the complexity of the input for the upcoming layers. Two types of pooling operation are there, max pooling and mean pooling. In max pooling, it gives the maximum value within the sub-matrix. The max-pooling is the most commonly used pooling type. In the case of mean pooling, the mean of all values is calculated from the sub-matrix and returns only one mean value (See in the figure 1) [4], [6], [7], [10].
Sometimes using the pooling layers can destroy the performance of the CNN model. Because the pooling layer helps us to discover the specific feature whether it is present in the layer or not irrespective of the location of that feature. So in that case our model can leave or forget the most valuable detail of the image [9].
For Example:



1	1	1	3
4	6	4	8
30	0	1	5
0	2	2	4

3	4
8	3

6	8
30	5

Figure 3: Working of Mean pooling and Max pooling [7]

H. Fully Connected Layer

It is the final layer of any CNN architecture in which every neuron is connected to every other neuron from the last layer before it. This layer is responsible for the classification. The fully connected layer can consist of a large number

DOI: 10.17148/IJARCCE.2022.11159

of parameters which makes the process computationally more complex. By the dropout technique, this limitation of the final layer can be avoided [8], [9].

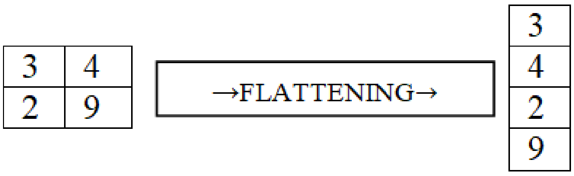


Figure 4: Working of a Flattening layer

II. Applications of CNN in Medical e-Diagnosis

A. CNN based Breast Cancer Classification (BCC)

Breast cancer is a very common disease in women and it is very life-threatening as well. If we diagnose it in the early stage then it can be treated. But the machines can be programmed to detect the cancerous cell in breasts' medical images. In the proposed paper [10] a CNN method is proposed for classification for breast cancer tissues. The dataset used in this solution was potentially large about 275,000, 50x50 pixel RGB images. This paper uses various CNN architectures and will compare their results. The model (say model A) proposed has only two convolution layers. The first layer is with 32 kernel size and the second layer has a 64 kernel size. To compensate for overfitting the dropout layer is used. Flatten layer is used to flatten the image preparing for the next layer input. The batch size of 128 is used with 12 epochs. The model performs with a slight difference on the training set and validation set. The output layer uses the softmax activation function and all other layers except the output layer are using ReLu(Rectified Linear Unit). The accuracy found for the model was 59% which is not enough to use this model for real-world applications. There is a huge requirement for potential improvements in this model.

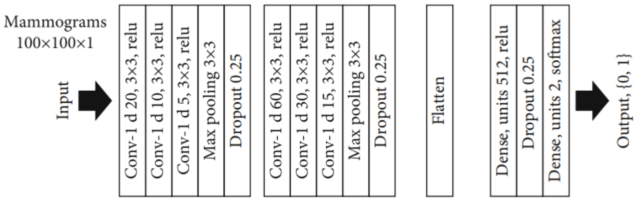


Figure 5: Network Architecture, Source: [10]

In [11] 6 convolution layers are used instead of just 2 as we saw in [10] as a result it increases the number of features for better pattern recognition. The 4 dropout layers are used to overcome the problem of overfitting. The accuracy of the model (say model B) this time increased to 76% which is a fair jump from the previous accuracy.

In [12] the model (say model C) proposed and implemented outputs an accuracy of 87% which is the best. In this, the five layers of CNN is implemented to classify the images.

Overall, we are getting a good accuracy of 87% with a Lil bit deeper approach in model C. Model C is capable to reduce the human mistakes in detecting breast cancer as well as it will decrease the cost of a breast cancer diagnosis. If this approach is implemented on a large scale then it can assist healthcare professionals to overcome some workloads. In developing nations the lack of doctors is a major issue that can be challenged by e-diagnosis.

B. CNN for Heartbeat Classification

Electrocardiogram (ECG) is a standard medical test that is used to monitor the electrical activity of the heart. It generates a rhythmic pattern of the heartbeat which can be useful to identify the normal and abnormal pattern of the working of the heart. Problems which used to generate in the heart affect the electrical activity of the heart. An ECG test can assist a healthcare professional to identify the reason for the problem in the heart. Risk of heart attack, irregularity in a heartbeat, problems with heart valves, blocked arteries and coronary artery disease etc. can be diagnosed by the ECG itself. The solution in [13] demonstrates the CNN model which consists of 9 different deep CNN layers to categorize the different types of heartbeats. Generally, the ECG signals in the output are found to be corrupted or attenuated with noise. Before feeding the dataset into the model it is pre-processed to remove the noise from the dataset. [14] presents a procedure to denoising the ECG signal while preserving the peak of the ECG signal. The psychological information of the patient is stored in the peak of the ECG signal that's why it is important to save the

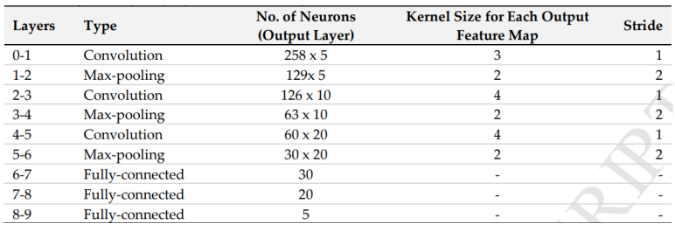
ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⪤ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11159

peak of the signal while denoising it. In [brij], a basis function called Daubechies mother wavelet is applied to the noisy ECG signal to reconstruct it into a noise-free signal also securing its peak while processing the signal. A summary of each layers are illustrated in Figure 6.
The database used is open source and extracted from the PhysioBook MIT-BIH Arrhythmia database. This database contains raw ECG signals that will undergo pre-processing for cleaning and denoising. The model proposed in this paper has 9 layers in the architecture which consists of three different types of layers convolution (Conv2D) layer, max-pooling (MaxPool2D) layer and fully connected (FC) layer (See figure 4). To achieve the optimal performance of the model the hyperparameters are carefully set. The regularisation value is 0.2, the learning rate is 3x10^{-3}, and the momentum is set to 0.7. These values of the parameter are selected therefore they can avoid overfitting and can control the learning rate during the training.



Layers	Type	No. of Neurons (Output Layer)	Kernel Size for Each Output Feature Map	Stride
0-1	Convolution	258 x 5	3	1
1-2	Max-pooling	129x 5	2	2
2-3	Convolution	126 x 10	4	1
3-4	Max-pooling	63 x 10	2	2
4-5	Convolution	60 x 20	4	1
5-6	Max-pooling	30 x 20	2	2
6-7	Fully-connected	30	-	-
7-8	Fully-connected	20	-	-
8-9	Fully-connected	5	-	-

Figure 6: A summary of the model used in [13]

This model feed with two different sets of data i.e. Set A and Set B. On set A, the accuracy was 91.64% and on set B, the accuracy was 91.54%. The implementation proposed in this paper is computationally expensive and could take several hours on GPU to train the model only. These types of implementation always need a large amount of data for better accuracy which can't be available every time. However, it makes the process of diagnosis from ECG signal fully automatic and which is unconcerned with the ECG signal noises. Overall, this implementation can save a potential amount of time for a healthcare professional and diagnosis of cardiovascular disease can be done over the cloud which will make healthcare more cost-effective.

C. CNN for Interstitial Lung Disease (ILD)

Interstitial lung disease (ILD) is related to the lung tissues. It affects breathing and the supply of oxygen into the bloodstream. It is almost irreversible so the lung transplant is the only option in case of ILD. But it can be diagnosed with deep learning with the help of High resolution computed tomography (HRCT) images of lungs. A customized CNN framework in [15] can diagnose it. This framework can learn the complicated features from the lung images. The architecture of the CNN framework contains only five layers i.e. a convolution layer, a Max-Pooling layer and three fully connected layers. The kernel of 7x7 and the output of 16 channels is applied to the first layer of the network. Max pooling kernel size is set to 2x2. The other three layers are fully connected layers with only 5 neurons in the last layer. The input images from the dataset are grayscale and texture-based so that there are not too many features present to learn so in this network architecture only one convolution layer is used. The activation function used in the convolution laser is ReLU (Rectified Linear unit) which increase the accuracy by 2.5%. The dataset used, publicly available, contains 113 sets of HRCT images with 2062 2D ROI marked demonstrating the ILD images. Five different categories were used to classify the ILD images i.e. normal (N), emphysema (E), ground glass (G), fibrosis (F), and micronodules (M). The dataset is divided into 10 parts and 9 parts are used for training and 1 part is used for validation purposes.

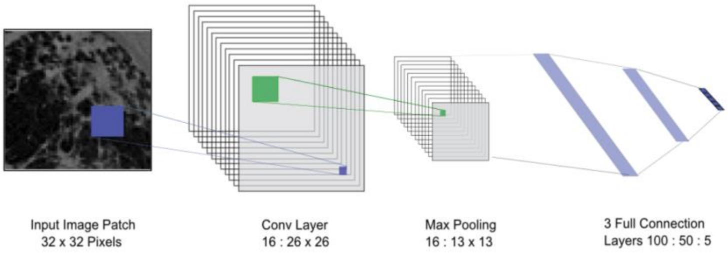


Figure 7: Customized CNN architecture for ILD, Source: [15]

The classification results that this network of CNN can classify complex features automatically hence no requirement for manual intervention. However, the major limitation was the small size of the dataset. The dataset required is quite large the model can predict and extract the features from training images.

III. CONCLUSION

In this work, we have illustrated that how CNN works and the application of CNN in medical e-Diagnosis. We stated in this review paper the importance of machines and computation in medical diagnosis. Lack of healthcare professionals and too much workload on them is a real-world problem that has been kept in mind while writing this review paper. So, in this paper, we discussed a deep learning concept and also the current progress in the field of medical e-Diagnosis through deep learning. Training a model with ANN can be more computationally expensive but the CNN is less expensive than the ANN and widely used deep learning technique for processing images, videos, and audios. According to the work done the accuracy, in breast cancer classification and heartbeat classification, was 87% and approx 91% respectively moreover in case of ILD diagnosis the results was the model trained was able to classify the complex features of ILD although the dataset available was very small in terms of size. After analyzing some of the cases of multilayer perceptron model for medical e-Diagnosis we came across a limitation of unavailability of the adequate size of dataset however the data augmentation helped to some extent. But always before training a model the size of the dataset should be taken care of by the researcher to result the acceptable accuracy.

IV. REFERENCES

[1] A. Rosebrock, “Deep Learning for Computer Vision with Python - Starter Bundle,” *Pyimagesearch*, p. 330, 2017.

[2] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[3] L. Deng and D. Yu, “Deep learning: Methods and applications,” *Found. Trends Signal Process.*, vol. 7, no. 3–4, pp. 197–387, 2013, doi: 10.1561/2000000039.

[4] K. O’Shea and R. Nash, “An Introduction to Convolutional Neural Networks,” pp. 1–11, 2015, [Online]. Available: http://arxiv.org/abs/1511.08458.

[5] J. Wu, “Introduction to Convolutional Neural Networks,” *Introd. to Convolutional Neural Networks*, pp. 1–31, 2017, [Online]. Available: https://web.archive.org/web/20180928011532/https://cs.nju.edu.cn/wujx/teaching/15_CNN.pdf.

[6] 3 & Richard Kinh Gian Do2 & Kaori Togashi1 Rikiya Yamashita1,2 & Mizuho Nishio1, “Convolutional neural networks: an overview and application in radiology https://doi.org/10.1007/s13244-018-0639-9,” *Springer*, vol. 195, pp. 21–30, 2018.

[7] P. Kim, “MATLAB Deep Learning,” *MATLAB Deep Learn.*, no. November 2013, pp. 121–147, 2017, doi: 10.1007/978-1-4842-2845-6.

[8] S. Albawi, T. A. M. Mohammed, and S. Alzawi, “Layers of a Convolutional Neural Network,” *Ieee*, p. 16, 2017.

[9] A. Ghosh, A. Sufian, F. Sultana, A. Chakrabarti, and D. De, *Fundamental concepts of convolutional neural network*, vol. 172. 2019.

[10] S. Z. Ramadan, “Using convolutional neural network with cheat sheet and data augmentation to detect breast cancer in mammograms,” *Comput. Math. Methods Med.*, vol. 2020, 2020, doi: 10.1155/2020/9523404.

[11] M. Nawaz, A. A. Sewissy, and T. H. A. Soliman, “Multi-class breast cancer classification using deep learning convolutional neural network,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 6, pp. 316–322, 2018, doi: 10.14569/IJACSA.2018.090645.

[12] Y. Zhang *et al.*, “Automatic Detection and Segmentation of Breast Cancer on MRI Using Mask R-CNN Trained on Non–Fat-Sat Images and Tested on Fat-Sat Images,” *Acad. Radiol.*, 2020, doi: 10.1016/j.acra.2020.12.001.

[13] U. R. Acharya *et al.*, “A deep convolutional neural network model to classify heartbeats,” *Comput. Biol. Med.*, vol. 89, pp. 389–396, 2017, doi: 10.1016/j.compbiomed.2017.08.022.

[14] B. N. Singh and A. K. Tiwari, “Optimal selection of wavelet basis function applied to ECG signal denoising,” *Digit. Signal Process. A Rev. J.*, vol. 16, no. 3, pp. 275–287, 2006, doi: 10.1016/j.dsp.2005.12.003.

[15] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, “Medical image classification with convolutional neural network,” *2014 13th Int. Conf. Control Autom. Robot. Vision, ICARCV 2014*, vol. 2014, no. December, pp. 844–848, 2014, doi: 10.1109/ICARCV.2014.7064414.