Праці НТШ Медичні науки
2021, Том 65, № 2
ISSN 2708-8634 (print) www.mspsss.org.ua

Proc Shevchenko Sci Soc Med Sci
2021, Vol. 65, 2
www.mspsss.org.ua ISSN 1563-3950

# Поглиблений аналіз основ технології глибинного навчання

Іван Волянський*

*Меморіальний онкологічний центр імені Слоуна-Кеттерінга, Відділення медичної фізики, м. Нью-Йорк, Нью-Йорк, США*

Глибинне навчання – це тип машинного навчання (МН), важливість якого в медичній сфері постійно зростає. В обчисленні різних показників воно часто працює краще, ніж традиційні моделі машинного навчання, і може справлятися з нелінійними функціями за допомогою функцій активації. Функції активації – це різні нелінійні функції, які використовують для обмеження значень, що поширюються на інтервал. У випадку глибинного навчання інформація поширюється далі, проходячи через різні шари зв'язків і функцій активації, перш ніж дійти до останнього шару. Після цього оцінюється функція втрат і поширюється назад через мережу для коригування зв'язків.

Згорткова нейронна мережа (CNN) – це форма глибинного навчання, яка насамперед використовується в обробці візуальної інформації. Мережі CNN працюють дуже добре за даними в таблиці, бо вони добре навчаються формам. CNN обчислюють скалярний добуток між шарами та ядрами в згортковому шарі перед об'єднанням, виводячи підсумкову статистику. Мережі CNN ліпші, ніж звичні нейронні мережі для цілей візуальної обробки інформації з огляду на низку причин, зокрема через розріджену взаємодію та еквіваріантність тлумачення.

***Ключові слова:*** глибинне навчання, штучний інтелект, машинне навчання, нейронна мережа.

OPEN ACCESS

**DOI 10.25040/ntsh2021.02.23**

**Адреса для листування**: Відділення медичної фізики, Меморіальний онкологічний центр ім. Слоуна-Кеттерінга, 1275 Йорк Авеню, м. Нью-Йорк, Нью-Йорк, 10065-6007

**E-пошта:** iawolansky@gmail.com

**Надійшла до редакції**: 01.10.2021

**Прийнята до друку**: 01.11.2021

**Опублікована онлайн:** 29.11.2021



**ORCID iD:**
Іван Волянський:
https://orcid.org/0000-0002-2109-3942

**Конфлікт інтересів**: автор декларує відсутність конфлікту інтересів.

**Дозвіл комісії з біоетики:** для цього огляду не потрібний.

**Фінансування**: не потрібне.

* Оприлюднено під час 5-го міжнародного симпозіуму «SMART LION». Перспективи в медицині та глобальне здоров'я, 7-9 жовтня 2021 року.

# A Deep Dive into the Basics of Deep Learning

Ivan Wolansky**

*Memorial Sloan Kettering Cancer Center, Department of Medical Physics, New York, New York, USA*

OPEN ACCESS

**DOI 10.25040/ntsh2021.02.23**

**For correspondence:** Department of Medical Physics, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, New York, 10065-6007

**E-mail:** iawolansky@gmail.com

**Received:** Oct, 1, 2021

**Accepted:** Nov, 1, 2021

**Published online:** Dec, 29, 2021



**ORCID:**
Ivan Wolansky:
https://orcid.org/0000-0002-2109-3942

**Disclosures:** the author declared no conflict of interest.

**Ethical approval:** Not required for this review.

**Funding:** Not applicable.

Deep learning is a type of machine learning (ML) that is growing in importance in the medical field. It can often perform better than traditional ML models on different metrics, and it can handle non-linear problems due to activation functions. Activation functions are different non-linear functions that are used to restrict the values propagated to an interval. In deep learning, information propagates forward, passing through different layers of weights and activation functions, before reaching the final layer. Then a cost function is evaluated and propagated back through the network to adjust weights.

A convolutional neural network (CNN) is a form of deep learning that is used primarily in imaging. CNNs perform significantly well with grid-like inputs because they learn shapes well. CNNs compute dot products between layers and kernels in a convolutional layer, prior to pooling, which outputs summary statistics. CNNs are better than trivial neural networks for imaging due to a number of reasons, like sparse interaction and equivariance of translation.

***Keywords:*** Deep learning, artificial intelligence, machine learning, neural network.

** Presented during 5th International Symposium "SMART LION". Medical Imagining and Global Health, October 7–9, 2021.

## Introduction

Deep learning is a form of machine learning (ML) that was inspired by the neuronal connections in the brain. In deep learning, there are neural networks with tens to hundreds of layers of neurons. Neural networks are represented by a hierarchy of layers. There is an input layer, then hidden layers that receive data from previous layers, and finally an output layer. As information propagates from the input layer, through hidden layers, and finally to the output layer, features and representations get understood with deeper and deeper complexity.

Deep learning has found increasing importance in medical research, particularly in imaging. Currently, the research includes models that can diagnose different diseases based on imaging, auto-segmentation algorithms that can draw around lesions, models that can determine lesion risk level, and more. As this research progresses, we will see more and more use of these kinds of models in clinical settings.

### *Why Deep Learning?*

Deep Learning generally performs better on important metrics when compared to traditional ML algorithms. However, with this comes the caveat that, in order to function well, deep learning requires a significantly larger amount of data. This, of course, can become problematic in medicine, where a lot of data sets are relatively small. In addition to this, it is important to note that deep learning requires a large amount of computational power when compared to older machine learning algorithms.

Another utility of deep learning is its ability to easily handle non-linear problems. This is due to activation functions. Additionally, deep learning has great utility because it does not rely on human-created features. Traditionally, in ML, features are selected or created by humans. This selection is based either on their impact on performance or on how much they explain data set variance. However, in deep learning, there is no need for human input because the model actually learns features on its own. This removes the subjectivity of feature extraction in the pre-processing step. This is called feature learning.

### *How Does it Work?*

A lot of people describe deep learning as being a sort of "black box" where we do not know what is happening. This is not entirely true. Outputs of these models are fairly straightforward. Cancer/no cancer and different risk levels are two examples of understandable outputs. However, it can be difficult to understand the hidden layers' nodes and how they justify results (known as explainability). Despite this, we are still capable of looking inside the models to see what each component is doing mathematically. Given this, at the scale of current models, this can be quite arduous, as there are models with millions to even billions of weights and computations. We can, however, still learn the basics of what happens in these models.

From a biological perspective, the input layer is similar to when your eyes see an image. This image gets transmitted through the neurons of the brain so that it can be understood.

Similarly, in deep learning, the input layer takes the numerical representation and propagates them through the network. Then, each node has the weight that connects to a node in the next layer. This being said, each weight is a numerical value. Finally, computation occurs and an activation function is applied prior to the information being propagated onwards. This is called forward propagation (Fig. 1).

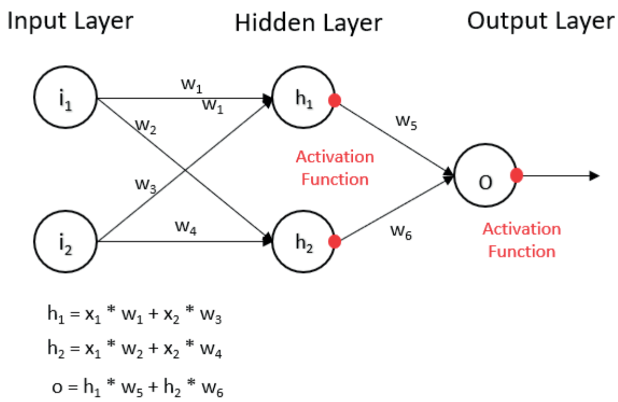


Figure 1. This Figure demonstrates a trivial 3-layer, 5-node neural network schematic

From this, each neuron knows to pass the information along or not (just like in biology, where we have an action potential for neurons).

It is important to note that just as in the case of explainability, there is a similar concept called interpretability. Essentially, this is how able we are to determine cause and effect based on the machine learning model.

***Activation Functions***

Activation functions are non-linear functions. This means that they do not follow the linear form $y = \beta_0 + \beta_1 x$. Their purpose is to transform and restrict the sum to an interval, which helps determine whether the information is passed on.

If the values were not restricted, then as they propagated through the network, they would keep getting larger as they pass from layer to layer. Additionally, it is important to note that failing to use non-linear activation functions over multiple layers simplifies the model mathematically to just having one layer because simply summing up layers is a linear process.

For example, the softmax activation function can take on multiple different values from a layer. It transforms and restricts them to an interval [0, 1] where they become probabilities. For example, the sum of 4 transformed values equals 1 (if you have four possible outcomes, each with a different chance of occurring, then their sum is 100%, or 1).

The network then chooses the transformed value with the highest probability and uses this as its output. This is the reason why the softmax activation function actually comes at the end of a neural network.

***The Most Popular Activation Functions***

The following images show two of the most popular activation functions: hyperbolic tangent (TanH) and rectified linear unit (ReLU).

In Figure 2, we see TanH. Here, the output is bounded between -1 and 1. Negative inputs (x) approach -1 (y), while positive inputs (x) approach 1 (y).

Finally, Figure 3 features ReLU. Here, the output (y) is 0 whenever the input (x) is less than 0. When the input (x) is greater than or equal to 0, the output (y) is that number because the function becomes linear (x = 2 mean y = 2, x = 100 means y = 100, etc.)

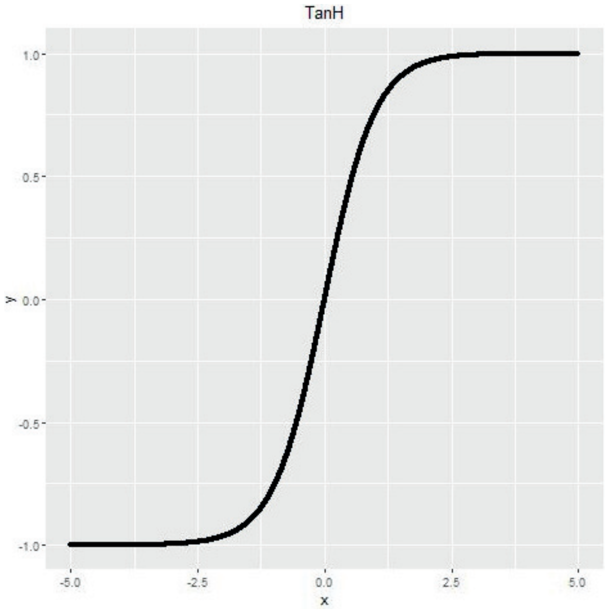


Figure 2. The figure portrays the activation function known as hyperbolic tangent (TanH)

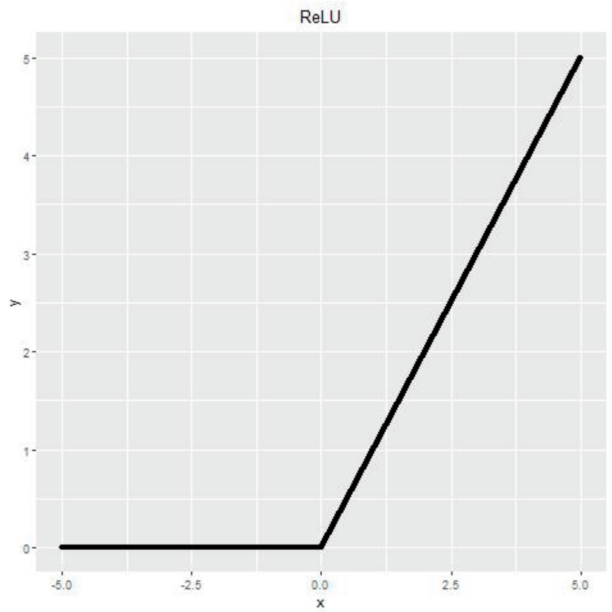


Figure 3. The figure portrays the activation function known as a rectified linear unit (ReLU)

***Cost Functions and Backpropagation***

Each model has a cost function. There are many different kinds of cost functions that can be used.

A cost function called "Mean Squared Error," which is primarily used in regression problems is:

$$c = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \qquad (1)$$

where $y_i$ is the actual value, and $\hat{y}_i$ is the predicted value. We want to minimize this function, as this means that the actual value and predicted value are closer together.

Based on the result of this function, the weights between nodes are changed through propagating them back. In essence, we are trying to change the weights so that the actual value and predicted value are even closer together, thus minimizing the cost function. This is called backpropagation.

We change the weights by finding weights closer to the optimal weights. This is done by using the stochastic gradient descent algorithm (SGD).

***Putting It Together***

Initialize weights randomly as values close to 0

Input your first observation into the input layer

Forward propagation – the neurons are activated and each activation is restricted by weights and activation functions. This continues until the output layer is reached.

The cost function is evaluated

Backpropagation – Weights get updated based on how much they impact the cost function. The user inputs the learning rate (part of the SGD optimization), and this determines how much the weights are updated.

When the whole training set has passed, this is called an epoch. You then would repeat more epochs.

A lower learning rate means the model "learns" more slowly, but also means that it is more precise in finding the best solution. A lower learning rate means a higher number of epochs, which in turn increases the time needed for model training.

What Form Do We Typically Use in Imaging?

The best type of deep learning model for medical imaging is typically considered the convolutional neural network (CNN). These models are particularly good with grid-like inputs, like a pixel or voxel matrix. This is because CNNs tend to learn shapes (tumor or not tumor) regardless of their location in an image. For example, if there are 2 different images but one has a tumor in the top right corner and the other has the tumor in the center of the image, the CNN will perform well regardless. In a traditional, trivial, deep learning model, the tumor would need to be in the same relative area to perform well.

***What is a CNN and how does it work?***

CNN takes an image matrix and then computes dot products in what is called a convolution layer. The dot product is between the data and a matrix called a kernel. In some sense, this can be thought of as how much the input layer or previous layer's nodes "agree" with one another. The output of this is a new, smaller matrix. This is called convolution.

Simply put, after convolution, like in trivial deep learning, an activation function is applied. Next, pooling is applied. Finally, there is a fully connected layer at the end.

***Pooling layers***

Pooling layers can sometimes replace the standard output of certain layers. The output summary statistics of the output from a layer decreases the number of computations and weights.

A popular pooling layer is max pooling, which will output max values in different areas of the matrix, depending on the filter size (another user input) of the pooling layer.

***Why use CNN?***

Unlike trivial deep learning which computes interactions between each input and output neuron, CNN's use of a kernel smaller than the input means that we can glean important information from hundreds or thousands of pixels at a time, instead of each individual pixel. This lowers the number of parameters which increases statistical efficiency and decreases computation time. This is called sparse interaction.

Additionally, the decrease in the number of parameters means that there are fewer

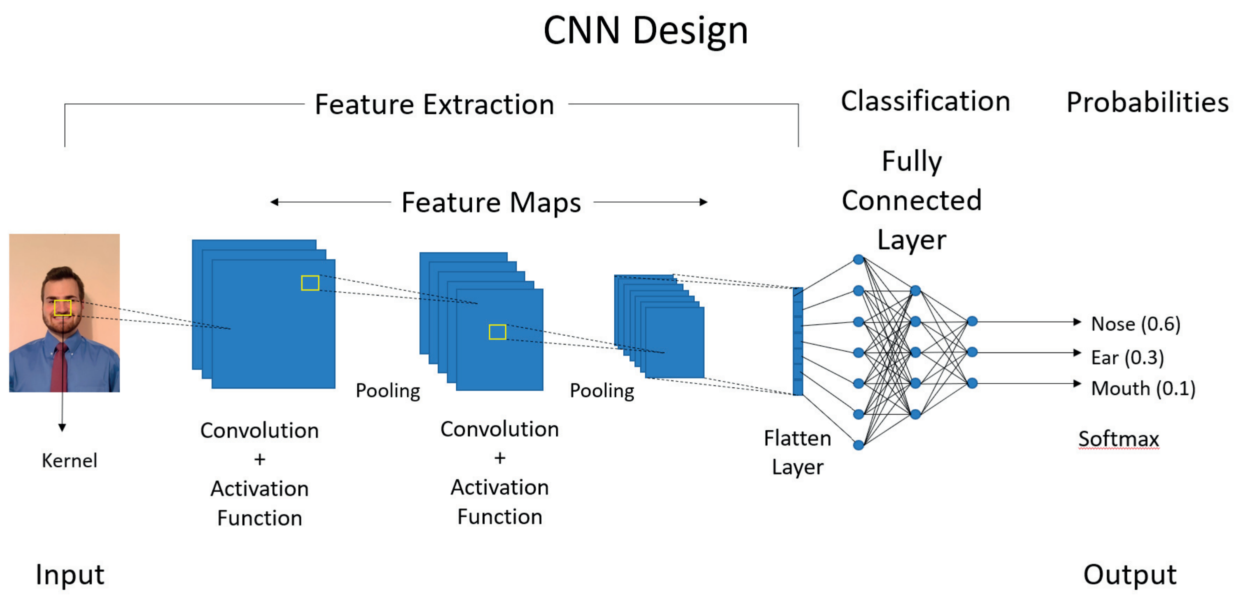


Figure 4. The figure demonstrates a convolutional neural network (CNN) scheme for recognizing facial structures.

chances of overfitting your model, which is when your model is trained so specifically to your data that it cannot be generalized and applied to other cases with new data.

Another reason to use CNN – it is fair to assume that computing a feature at one point could probably have useful information at another point (grass is still grass, for instance). A trivial deep learning model would only use the weights once, but a CNN shares parameters, so weights applied in one place are the same elsewhere.

Finally, this parameter sharing means that CNNs have the equivariance of translation. This is when changing the input in some way would affect the output. More specifically, this means that the position of an object in the input image does not need to be fixed for the CNN to detect it.

***Putting CNN Together***

To sum up, one of the technologies with growing importance in medicine is deep learning. In particular, a form of deep learning called the convolutional neural network is relevant in the case of imaging research. This technology has the potential to aid physicians clinically in diagnosis, prognosis, and treatment.

## References

1. https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939
2. https://www.mathworks.com/videos/introduction-to-deep-learning-what-are-convolutional-neural-networks--1489512765771.html
3. https://mlfromscratch.com/neural-networks-explained/#/
4. https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/
5. https://towardsdatascience.com/everything-you-need-to-know-about-activation-functions-in-deep-learning-models-84ba9f82c253
6. https://towardsdatascience.com/what-is-deep-learning-and-how-does-it-work-2ce44bb692ac
7. https://machinelearningmastery.com/what-is-deep-learning/
8. https://towardsdatascience.com/deep-learning-basics-a-crash-course-3213aa9e477c
9. https://towardsdatascience.com/machine-learning-for-beginners-an-introduction-to-neural-networks-d49f22d238f9