**American Journal of Engineering Research (AJER)**
e-ISSN: 2320-0847 p-ISSN: 2320-0936
Volume-9, Issue-1, pp-113-118
www.ajer.org
Research Paper
Open Access

# Deep Learning with a particular focus on Architectures, Application and Recent Trends

Mensah Solomon[1], Wei Song[1]
*1School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China*
*Corresponding Author: Mensah Solomon*

***ABSTRACT:*** *Deep learning has become one of the most popular scientific research trends in our contemporary society. Deep learning has taken center stage in our technological world through its revolutionary advancement in computer vison, natural language processing and machine learning among others which is deeply rooted in artificial intelligence. In a digital era where the use of data has become invaluable, deep learning has the ability to make better use of datasets for feature extraction. Due to the rapid evolution of deep learning techniques, new developments and significant breakthroughs have been made in domains such as medical research, industrial automation, electronics and governments among others for solving the society's day-to-day problems. This is seen in stock market analysis, speech and text recognition, cancer detection, adaptive testing, biological imaging etc. In this paper, deep learning will be evaluated from the perspective of its recent trends, deep learning architectures, Applications and techniques of this architectures and applicable areas of deep learning. Finally, the paper will discuss the practical use of deep learning, some applications prospects of such a nascent advancement in the world of computer science.*

***KEY WORDS:*** *Deep learning, artificial intelligence, machine learning, pattern recognition, Deep Learning Architectures*

---

Date of Submission: 15-01-2020 Date of acceptance: 31-01-2020

---

## I. INTRODUCTION

Owing to several researches conducted on artificial intelligence and machine learning, deep learning has invariably been an essential and extensively investigated subject matter. The field of machine learning and deep learning are subsumed under artificial intelligence which lays much emphasis on the development machines, thinking and working like humans. That is, Artificial Intelligence (AI) makes it possible for machines to not only learn from experiences and adjust to new inputs but also perform human-like tasks. . [1]The primary aim is to allow a machine to learn useful information just like humans do through approaches such as supervised learning, unsupervised learning and semi-supervised learning.Deep learning is a new area of machine learning which has gained popularity in recent past. Deep learning refers to the architectures which contain multiple hidden layers (deep networks) to learn different features with multiple levels of abstraction. [2] Deep learning algorithms seek to exploit the unknown structure in the input distribution in order to discover good representations, often at multiple levels, with higher level learned features defined in terms of lower level features.

Deep learning algorithms can learn the right set of features, and it does thisin a much better way than extracting these features using hand-coding. Instead ofhandcrafting a set of rules and algorithms to extract features from raw data, deeplearning involves learning these features automatically during the training process. In deep learning, a problem is realized in terms of hierarchy of concepts, witheach concept built on the top of the others. The lower layers of the model encodesome basic representation of the problem, whereas higher-level layers build uponthese lower layers to form more complex concepts onlyrecently that deep networkismade a big reappearance by achieving spectacular resultsin speech recognition and computer vision tasks. Deep learning works on the artificial neural network system (ANN). The neural activity in our brainsis far more complex than might be suggested by simply studying artificial neurons.The learning mechanisms used by deep learning models are in no way comparableto the human brain, but can be described as a mathematical framework for learningrepresentations from data.

## II. OVERVIEW ANTECEDENTS OF DEEP LEARNING

The history of deep learning can be dated back to 1943, when Walter Pitts and Warren McCulloch created a computer model based on the neural networks of the brain. They used a combination of algorithms and mathematics they called “threshold logic” to mimic the thought process. Over the years, deep learning technology has steadily evolved by making huge leaps in the world of computer science. In 1960, Henry J. Kelley came up with the basics of a Continuous Back Propagation Model. Two years on, Stuart Dreydus developed a simpler version on the basis of the chain rule. Even though the notion of back propagation (the backward propagation of errors for purposes of training) did exist in the early 1960s, it was clumsy and inefficient and would not become useful until 1985 [4].System for mapping and recognizing similar data). Sepp Hochreiter and Juergen Schmidhuber developed LSTM (long short-term memory) for recurrent neural networks in 1997.

## III. DEEP LEARNING ARCHITECTURES

This section will some of the popularly knownarchitectures applied in deep learning.

### 3.1. Convolutional Neural Network

Convolutional networks are the first examples of deep architectures that have successfully achieved a good generalization on visual inputs. [5,6] They are thebest known method for digit recognition. [7] They can be seen as biologicallyinspired architectures, imitating the processing of “simple” and “complex” corticalcells which respectively extract orientations information (similar to a Gaborfiltering) and compositions of these orientations.The main idea of convolutional networks is to combine local computations(convolution of the signal with weight sharing units) and pooling. The convolutionsare intended to give translation invariance to the system, as the weightsdepend only on spatial separation and not on spatial position. Convolutional neural networks are powering core ofcomputer vision that has many applications, which include self-driving cars, robotics,and treatments for the visually impaired. Some of the CNNs architectures include; LeNet, AlexNet, ZFNet, GoogLNet and ResNet among others. Architecture of CNN

CNN is depicted in the net and contains eight layers with weights the first five are convolutional and the remaining three are fully connected. The output of the last fully connected layer is fed to a 1000-way soft max, which produces a distribution over the 1000 class labels. Our network maximizes the multinomial logistic regression objective, which is equivalent to maximizing the average across it cases of the log-probability of the correct label under the prediction to distribute. The kernels of the second, fourth, and fifth convolutional layers are connected only to those kernel maps in the previous layer which reside on the same GPU (see Figure 1). The kernels of the third convolutional layer are connected to all kernel maps in the second layer. The neurons in the fully connected layers are connected to all neurons in the previous layer. Response-normalization layers follow the first and second convolutional layers. Max-pooling layers, of the kind described in Section 3.4, follow both response-normalization layers as well as the fifth convolutional layer. The ReLU non-linearity is applied to the output of every convolutional and fully-connected layer. The first convolutional layer filters the 224×224×3 input image with 96 kernels of size 11×11×3 with a stride of 4 pixels (this is the distance between the receptive field centers of neighboring 3We cannot describe this network in detail due to space constraints, but it is specified precisely by the code and parameter files provided here:

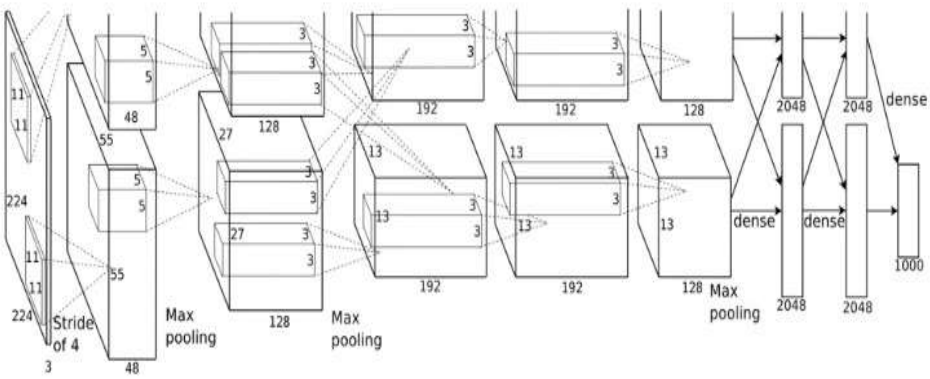


**Figure 1.** CONVOLUTIONAL NUERAL NETWORK

Figure1 is A CNN of GPUsshowing the delineation of responsibilities between two GPUs.one runs the layers parts at the top of the figure whiles the other runs the layers parts at the bottom .the GPUs connect only on a certain layers .if thenetwork inputs is 150 , 528 – dimensional ,and the number of neurons in the networks remaining layers isgiven by 253 ,440 – 186 ,624 -64-896-43,264-4096-4096-1000

The second convolutional layer takes as input the (response-normalized and pooled) output of the first convolutional layer and filters it with 256 kernels of size 5 × 5 × 48. The third, fourth, and fifth convolutional layers are connected to one another without any intervening pooling or normalization layers. The third convolutional layer has 384 kernels of size 3 × 3 × 256 connected to the (normalized, pooled) outputs of the second convolutional layer. The fourth convolutional layer has 384 kernels of size 3 × 3 × 192, and the fifth convolutional layer has 256 kernels of size 3 × 3 × 192. The fully connected layers have 4096 neurons each.

### 3.2. Autoencoder

Autoencoders (also called Autoassociators) are a family of neural networks for which the input layer is the same as the output layer, as well as an unsupervised learning algorithm. [8,9] They work by compressing the input into a latent-space representation, and then reconstructing the output from this representation. In more terms, autoencoding is a data compression algorithm where the compression and decompression functions are data-specific, lossy and learn automatically from examples. They have been used as building blocks to build a deep multi-layer neural network as well as reducing the dimensionality of the data. [10,11] An autoencoder takes a set of typically unlabeled inputs, and after encoding them, tries to reconstruct them as accurately as possible. As a result of this, the network must decide which of the data features are the most important, essentially acting as a feature extraction engine.

### 3.3 Deep Boltzmann Machine

The deep Boltzmann machinecontains many layers of hidden variables, and has no connections between the variables within the same layer. This is a special case of the general Boltzmann machine (BM), which is a network of symmetrically connected units that make stochastic decisions about whether to be on or off. While having very simple learning algorithm, the general BMs are very complex to study and very slow to compute in learning. In a DBM, each layer captures complicated, higher-order correlations between the activities of hidden features in the layer below. DBMs have the potential of learning internal representations that become increasingly complex, highly desirable for solving object and speech recognition problems.

### 3.4. Recurrent Neutral Network

Recurrent neural networks (RNNs) can be regarded as a class of deep generative architectures when they are used to model and generate sequential data. [14] The “depth” of an RNN can be as large as the length of the input data sequence. RNNs are very powerful for modeling sequence data (e.g., speech or text), but until recently they had not been widely used partly because they are extremely difficult to train properly due to the well-known “vanishing gradient” problem. Recent advances in Hessian-free optimization have partially overcome this difficulty using second-order information or stochastic curvature estimates [15]. In the recent work of Martens and Sutskever, RNNs that are trained with Hessian-free optimization are used as a generative deep architecture in the character-level language modeling tasks, where gated connections are introduced to allow the current input characters to predict the transition from one latent state vector to the next [16]. Such generative RNN models are demonstrated to be well capable of generating sequential text characters. More recently, substantive research has explored new optimization methods in training generative RNNs that modify stochastic gradient descent and show these modifications can outperform Hessian-free optimization methods [17, 18].

## IV. DEEP BELIEF NETWORK

Deep belief network (DBN): probabilistic generative models composed of multiple layers of stochastic, hidden variables. The top two layers have undirected, symmetric connections between them. The lower layers receive top-down, directed connections from the layer above. DBNs are learned one layer at a time by treating the values of the latent variables in one layer, when they are being inferred from data, as the data for training the next layer. This efficient, greedy learning can be followed by, or combined with, other learning procedures that fine-tune all of the weights to improve the generative or discriminative performance of the full network.

### 4.1 Restricted Boltzmann Machine

Restricted Boltzmann machine (RBM): a special BM consisting of a layer of visible units and a layer of hidden units with no visible-visible or hidden-hidden connections. The RBM discussed above is a generative model, which characterizes the input data distribution using hidden variables and there is no label information involved. However, when the label information is available, it can be used together with the data to form the joint "data" set. Then the same CD learning can be applied to optimize the approximate "generative" objective function related to data likelihood. Further, and more interestingly, a "discriminative" objective function can be defined in terms of conditional likelihood of labels. This discriminative RBM can be used to "fine tune" RBM for classification tasks [19].

**4.3. Deep Stacking Network**

Deep Stacking Networks (DSN) is also acknowledged as deep convex networks. DSN is different from other traditional deep learning structures. It is called deep because of the fact that it contains a large number of deep individual networks where each network has its own hidden layers. The DSN believes that training is not a particular and isolated problem, but it holds the combination of individual training problems. The DSN is made up of a combination of modules which are part of the network and present in the architecture. There are three modules that work for the DSN. Here every module in the model is having an input zone, a single hidden zone and an output zone. Subroutines are placed one over the top of another with the input to every module is taken as the outputs of the preceding layer and the authentic input vector[20].

**4.4. Recursive Neural Network**

Recursive Neural Networks, like Recurrent Neural Networks, can deal with variable length input. The primary difference is that Recurrent Neural Networks have the ability to model the hierarchical structures in the training dataset. Images commonly have a scene composed of many objects. Deconstructing scenes is often a problem domain of interest yet is nontrivial. The recursive nature of this deconstruction challenges us to not only identify the objects in the scene, but also how the objects relate to form the scene. A Recursive Neural Network architecture is composed of a shared-weight matrix and a binary tree structure that allows the recursive network to learn varying sequences of words or parts of an image.

**4.5. Deep Long-Short Term Memory Neural Network (LSTM)**

LSTM networks are the most commonly used variation of Recurrent Neural Networks. LSTM networks were introduced in 1997 by Hochreiter and Schmidhuber. The critical component of the LSTM is the memory cell and the gates (including the forget gate, but also the input gate). The contents of the memory cell are modulated by the input gates and forget gates. Assuming that both of these gates are closed, the contents of the memory cell will remain unmodified between one time-step and thenext. The gating structure allows information to be retained across many time-steps, and consequently also allows gradients to flow across many time-steps. This allows the LSTM model to overcome the vanishing gradient problem that occurs with most Recurrent Neural Network models.

## V. APPLICATION OF DEEP LEARNING IN MARKETING

A car online dealership and we want to use real-time bidding (RTB) as a mechanism to buy and space for our product on other websites -- for retargeting purposes.

RTB is an automated process that takes place in a short time frame of under 100 milliseconds. When a user visits a website, an advertiser is alerted, and a series of actions determines whether or that advertiser bids for an ad display.

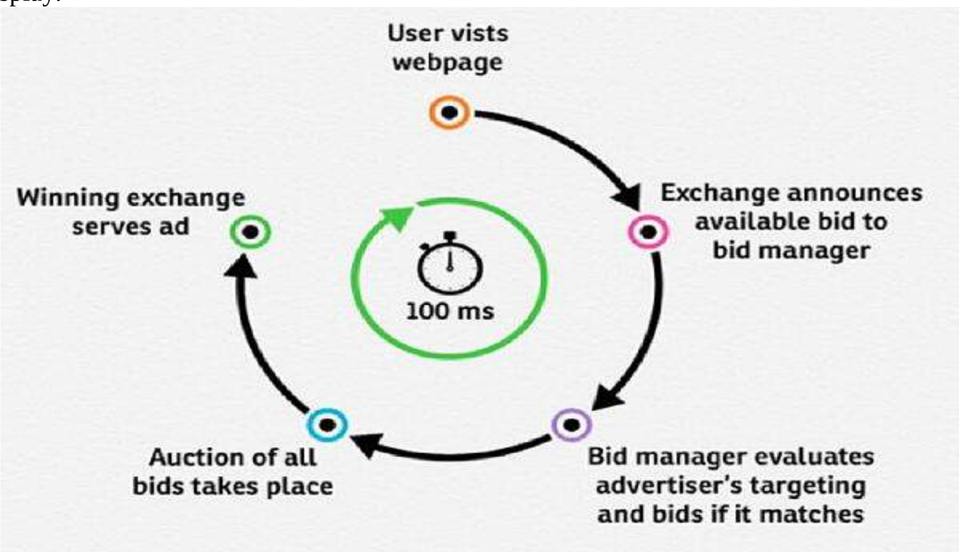


**Figure 2**. Indication auction sales with website and deep learning

RTB, we use software to decide if we want to bid for a certain product -- the software will make a decision by predicting how likely, the website visitor is to buy one of our products. We call that "buying propensity."
In this case, we will use deep learning to make this prediction. That means our RTB software will use a neural network to predict the buying propensity.

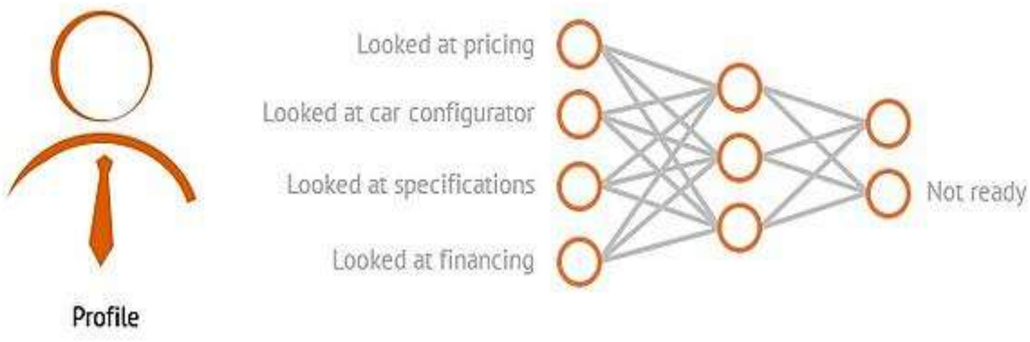


**Figure 3.** Neural network

The neural network inside our RTB software consists of neurons and the connections between them. The neural network on the above image has only a handful of neurons. A digital neural network have ever millions of neurons.

In this image, we want see if a website customer is likely to buy a product, and if we should pay an ad to target it. The result will depend on the interests and actions of the website customer.
To predict the buying propensity, we must choose features that are key to defining this person's digital behavior

- One Pricing, product, and specifications, money: The website customer is interested in the product, and "ready to buy." Conclusion: We should display on the product
- The website customer is not interested in the product, Conclusion: Do not show any product

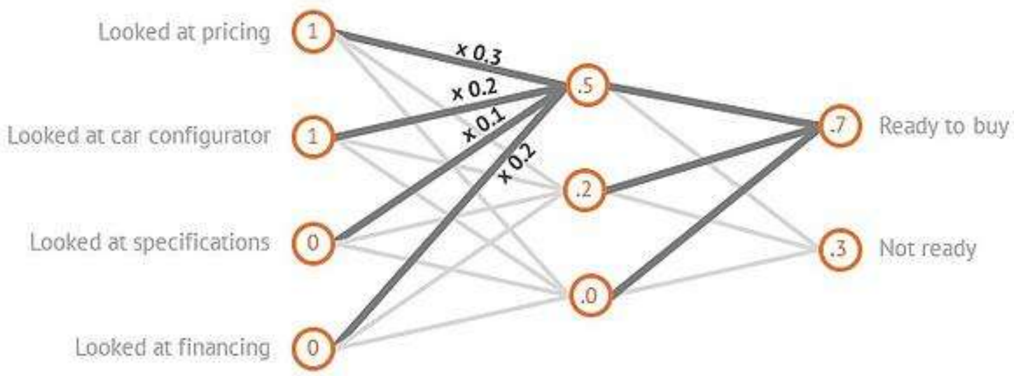


**Figure 4.** Functions of neural networks

In this Figure, a website customer looked at the Pricing a product but she skipped Specifications and Financing. Using the numerical system above, we get a "score" of a good score which means that there is chance this customer is "ready to buy" our product.

## VI. CONCLUSION

Applying the concepts of deep learning to drive automated training is made possible using these tools. While there are, other commercial solutions like that also available in the market

Applications of deep learning will become more present in today's time and achieve the results that can help drive many benefits for a business and ultimately.A recent report from FuturumResearch — commissioned by Automation Anywhere — discovered that more than half of North American businesses today have adopted intelligent automation.

These findings show that RPA is no longer an emerging technology but rather a mission-criticalcomponent for businesses, driving immense value due to its ability to automate labor-intensive tasks so that humans can refocus their efforts to work that drives impact and satifications

In this paper, we expatiate deep learning in the round, including the definition of deep leaning, the application and some deep learning arcthactures.

In addition, it is an important trend to use deep learning in engineering; we should make efforts on this.Scientists should pay attention to the new technique of deep learning and enlarge the application.

## REFERENCES


[1]. Patterson, J. & Gibson, A. Deep Learning: A Practitioner's Approach. United States of America: O'Reilly Media, 2017.
[2]. Goodfellow, I., Bengio, Y. &Courville, A. Deep Learning. Cambridge: MIT Press, 2016.
[3]. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature, 521, 436-444, 2015.
[4]. Foote, K. D. (2017). A Brief History of Deep Learning. Retrieved from https://www.dataversity.net/brief-history-deep-learning/#
[5]. LeCun Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. &Jackel, L. D. Handwritten digit recognition with a back-propagation network. In NIPS, 1990.
[6]. LeCun, Y., Bottou, L., Bengio, Y. &Haffner, P. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86:2278–2324, 1998.
[7]. Jarrett, K., Kavukcuoglu, K., Ranzato, M. &LeCun, Y. What is the best multi-stage architecture for object recognition? In ICCV, 2009.
[8]. YoshuaBengio, Learning Deep Architectures for AI, 2009, Dept. IRO, Universite de Montreal, Technical Report 1312.
[9]. P. Vincent, H. Larochelle Y. Bengio and P.A. Manzagol, Extracting and Composing Robust Features with DenoisingAutoencoders, 2008, Proceedings of the 25th International Conference on Machine Learning (ICML2008), pp. 1096-1103.
[10]. YoshuaBengio, Pascal Lamblin, Dan Popovici, and Hugo Larochelle, Greedy Layer-Wise Training of Deep Networks, 2007, Advances in neural information processing systems, pp. 153-160.
[11]. G. E. Hinton and R. R. Salakhutdinov, 2006, Reducing the dimensionality of data with neural networks, Science, 313(5786):504–507.
[12]. Salakhutdinov R. and Hinton, G. "A better way to pretrain deep Boltzmann machines," Proc. NIPS, 2012.
[13]. Srivastava, N. and Salakhutdinov R. "Multimodal learning with deep Boltzmann machines," Proc. NIPS, 2012.
[14]. Sutskever, I., Martens J., and Hinton, G. "Generating text with recurrent neural networks", Proc. ICML, 2011.
[15]. Martens J. "Deep learning with Hessian-free optimization," Proc. ICML, 2010.
[16]. Martens J. and Sutskever, I. "Learning recurrent neural networks with Hessian-free optimization," Proc. ICML, 2011.
[17]. Bengio, Y., Boulanger, N., and Pascanu. R. "Advances in optimizing recurrent networks," Proc. ICASSP, 2013.
[18]. Sutskever. I. "Training Recurrent Neural Networks," Ph.D. Thesis, University of Toronto, 2013.
[19]. Larochelle, H. and Bengio, Y. "Classification using discriminative restricted Boltzmann machines," Proc. ICML, 2008.
[20]. Dargan, S., Kumar, M., Ayyagari, M. R. & Kumar, G.A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning. Computational Methods in Engineering, 2019. https://doi.org/10.1007/s11831-019-09344-w


Mensah Solomon, et.al. "Deep Learning with a particular focus on Architectures, Application and Recent Trends". *American Journal of Engineering Research (AJER),* vol. 9(01), 2020, pp 113-118.