**IOSR Journal of Engineering (IOSRJEN)** **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 70-77

# Neural Networks

A.C. Sounthararaj, Dhanya.V. A, Deepika. G. Krishnan
*Assistance professor Department of BCA & MSc.SS Sri Krishna college of Arts and Science Kuniyamuthur, Coimbatore.*
*Department of BCA & MSc.SS Sri Krishna college of Arts and Science Kuniyamuthur, Coimbatore.*
*Department of BCA & MSc.SS Sri Krishna college of Arts and Science Kuniyamuthur, Coimbatore.*

***Abstract:*** *We all know how computers are evolved generation after generation since 1946. Starting with vacuum tubes or transistors science has modernized with evolution of micro processor and so on to make computers even better for computing. Now science is trying to bridge the gap between computing and thinking. This is possible with latest advancement of science under the field of artificial intelligence (AI). This has reduced the efforts of man and also has bridged the gap between computing and manual thinking. Now even more updated versions of AI have introduced the concept of "Neural Networks" which completely make machines to think and respond like human. This feature makes it different from other computer and also that no other latest computers are able to respond accordingly to queries. Neural networks are the concept of making computer works similar to human brain. The computation process is performed by implementation the way how Neurons are fired / triggered and stimulated in any human brain. At the course computation, large number of computers works simultaneously to produce result in more efficient accurate format. These technologies are able to solve complex hypothetical queries at an instance as like a human brain and respond in the same ways as human do.*

***Keywords:*** *neural networks, computing, artificial intelligence (AI)*

## I. Introduction

Neural networks, which are part of capital AI perform computation based on huge amount of collection of neural units. These neural units solve problem exactly as like a human brain. Generally each and every neural unit is connected to others them emphasizes the efficiency of the active state of connected neural units. These forms of computing commonly have two functions namely, summation function and threshold function. The summation function performs the task of combining all the input modules together. The threshold function also known as limiting function there are present in each and every connection as well as the unit itself must surpass before it reaches other neurons. These types of technologies have an advancement of self learning. These are designed in such a way that they respond accordingly to situation or quires other then explicitly specifying the program. This is the technology where it is trained rather than mastered by a program. Thus it can excel in all fields to solve the program and also in future deduction which is the milestone for traditional computer program. Generally a neural network has a cubic design with a multi layer, which has a single path traversal from front to back. Back propogation is done once when the forward stimulation is used to reset weight on the front neural units. This is also sometimes done with the combination of training the network about where the current result is known. The more advanced and modern net works are able to produce effective stimulation and inhabitation with connection interacting in much more chaotic and complex format. The most advanced once are the dynamic neural networks that are capable of new connections and new units dynamically by disabling the other connections of units.

**Goals and researches**

The main goal of the neural network is to solve problem exactly similar to the human brain. It aims on making the computers think rather than compute. The computational process performed in such a way that they are similar to that of thinking / thought process in human brain. It solves problem as human do. At present the modern neural networks are working on only with few thousands and millions of connections which is equivalent to the computation power of a worm. This must be stimulated as complex as human brain with introducing several more connection as marked at human brain. Researchers have recognized new patterns in neural networks. One of the latest approached in a method of establishing the connections with span [12]. This concept much further thus links processing layers rather than adjusting localized neurons. Another brain research recognizes the different signal of neurons at a particular time interval which is much more complex than the simple on or off signal. These are technologies based on real number system. Thus the core value of axon is represented using zero and one. The end result of these systems is generally unpredictable. Some of the systems after self learning are trained to become great problem solvers and some failed. The training undergoes

several thousand cycles of interaction typically occurred. Generally it is hard to solve the complex task such as computer vision and speech recognition with simple rule based programming. These tasks are accomplished only by the means of neural networks. Neural networks have been used for longer time from late 80s and marked direction shift from high level (symbolic) artificial intelligence to low level (sub-symbolic) machine learning. High level artificial intelligence are characterized by expert system with knowledge embodied in if-then rules. Whereas low level machine learning is characterized by knowledge embodied in the parameters of the dynamic system. Researchers are trying to make it much more efficient than the existing ones. It is trying to perform the task of thinking in computer rather than computing.

**Antic history**

In the year 1943 Warren Mc Colloch and Walter Pitts came up with an idea of computational models for neural networks which traces its origin from mathematical and algorithm called threshold logic. Due to these models neural networks were split in to two different and distinct approaches. One focused on biological process in the brain and the other on the application of the neural networks to artificial intelligence. A psychologist named Donneld Hebb in later 1940s evolved a new hypothesis of learning "Hebbian learning". In 1948 researches started implementing these ideas with Turing's B –type machines [13].

In 1954 FARLEY and WESLEY A.CLARK bought computational machines under use and called them to be calculators [14]. This was done to stimulate habbin learning at the laboratories of MIT. The other neural networks computational machines were developed by ROCHESTER, HOLLAND, and HABIT AND DUDA.

In 1958 FRANK ROSEN BLATT came up with a new algorithm for pattern recognition called 'PERCEPTRON'. This algorithm was based on 2 layer computer learning network using simple addition and subtraction [15]. RESENBLATT during his early models of basic perceptron did not describe circuits. With simple mathematical notations he implemented circuits using EX-OR (EXCLUSIVE OR). However this was not used until PAUL WERBOS created an algorithm called back propagation.

In 1969 MARVIN MINSKY and SEYNOR PARPERT was the two scientists who found the major issues with the computational machines [07]. One was Perceptrons were not capable of processing the EX-OR circuits. While the other was that the machine was not much efficient to handle the long run time requirements by large neural networks. Thus the researches were stagnated till these key issues were solved.

Back- propagation and resurgence: great turnover occurred once after the invention of back propogation algorithm [10]. These algorithms were able to solve the problem of EX-OR circuit efficiently and also they were quickly trained for multi layer neural networks in the year 1975.

The concept of connectionism became much popular in mid of 1980s. This was based on parallel distributed processing. DAVID E RUMELHART and JAMES MC CLELLAND clearly proved how connectionism was used in computers to stimulate neural process in their text book. Neural networks which are implied in AI was the simplified model of neural processing. There is no close traces of the neural net work model and biological architecture of the brain, thus neural net works are able to give the mirror reflection of brain only to smaller extent and not the full fledged version.

Support vector machines and other similar method such as liner classifiers eventually overtook neural networks in machine learning popularity. As GPU and distributed computing has improved the power of computing, the problem of image and visual recognition came in front of ANN [05]. (Artificial Neural Network). Thus neural networks were developed in large scale and have no become "deep learning". By emphasizing the use of model of parallel hardware implementation, deep learning which is re-branching of neural networks are being implied.

Up gradation from 2K06: Computational devices are generally made up of CNOS. For both Bio-physical stimulation and Neuromarphic computing. The latest advancement is to create nano devices for very large scale principle component analysis and convolution. If this turns up successfully, this would create a new era with the more advanced and new classes of computing [11]. This is because it emphasis on learning rather than programming, and also it is fundamentally an analog devise rather than the digital even though the first installation was done using CMOS Digital devise,

Research group of JURGEN SCHMIDHUBER at SWIZ AL LAB ID SIA has bagged process at eight international competitions in pattern recognition and machine learning during the years 2009 to 2012 based on recurrent neural network and deep feed forward neural networks. [08] One of the best examples was the bi-directional and multi directional long and short term memory (LSTM) of ALEX GRAVES ET AL one hatric prices at competitions for handwriting recognition

Any prior knowledge about the three different languages in the year 2009 at the international conference on document analysis and recognition (ICDAR).

DAN CIRESAN and his colleagues at IVSIA implemented GPU for computation and eventually won seven pattern recognition contests. Contest include IJCNN 2011 Traffic sign recognition, the ISBI 2012

Segmentation of neural challenges etc.. These neural networks are the first artificial pattern recognizes to achieve human competitiveness and also super human performance such as traffic sign recognition (IJCNN 2012) or (MNIST), Hand written Digital Problem of YANN LECUN at NYU.

"Standard Architecture of vision " which was inspired by the simple and complex cell identified by DAVID H,HUBEL and TORSTEN WIESEL similar to neocognitron by KUNIHIKO FUKUSHIMA in 1980. These machines can be pre-trained by unsupervised method of GEOFF TINTON'S lab at university of Toronto.

At 2012 contest, team from the lab which was sponsored by MERCK bagged the price. This team was held to design software to help find molecules that might lead to new drugs.

**MODELS:**

Generally neural networks that are implied in AI are referred as artificial neural networks (ANN). These are generally mathematical models defining a function or a distribution over both. [06] Sometimes models are associated with learning algorithm and learning rules.

***NET WORK FUNCTIONS:*** The term ANN refers to the interconnection between neurons at different layers of each system. Generally a complex system has more number of layers of neurons and sometimes some have increased layers of input and output neurons [4]. These neurons have a parameter called "weight" that manipulates the date in calculation. ANN defines three types of parameters. They are

1. The interconnection pattern between the different layers of neurons
2. The learning process for updating the weight of the interconnections
3. The activation function that converts the Neuron's weight input to its output activation.

Generally neural network functions are defined as a composition of other functions. This can be convenient ly represented as the net work structure with arrow debiting the dependency between the variables. A commenly and most widely used type of composition is the non linier weighted sum, where we use some of the predefined functions such as hyperbolic tangents.

In the above diagram the dependency between the variables are indicated by the arrows [6]. This can also be interpreted in two ways:

Functional view: In this first the input is transformed in to a three dimensional vector and them to two dimensional and finally to the target. [7] This is most commonly encounter in optimization.

Probabilistic view: In this method each proceeding random variable is dependent upon its successive random variable. This method is commonly encountered in graphical methods.

The two models view are largely equivalent. In either of the cases network architectural consideration tells us that the components of the individual layer are independent of each other. This eventually enables degree of parallelism in the implementation.

Previous versions of networks are called as feed-forward and they are directed acyclic graphs. But once when the networks are cyclic then it is called recreant. This shows the dependency on itself. However an implied temporal dependency is not shown.

***LEARNING:*** The learning is feature that attracts people towards ANN. When a particular task is assigned to find the solution, learning makes use of set of observation to in which solves the task in more optimized manner. The ost of optimized solution is the least of other solutions for the task. Cost function is generally important concept in learning. [3] It determines how for the solution is to that of the optimized solution for other problem or task that is to be solved. Learning algorithm is capable of performing a search and tells us the smallest possible cost. Once when the solution is dependent on some other data, then the cost must be a function of observations. [1] Otherwise no modeling is performed related to the data. It is often defined as statistics where only approximation is possible. The cost is generally minimized over a sample of data rather than a entire distribution generating the data. The online machine learning is mostly preferred for the classes where the distribution changes slowly overtime.

1. Choosing a cost function: while it is possible to define sum arbitrary ad-hoc cost function, frequently a particular cost will be used either because it has desirable properties or because it arises naturally from particular formulation of the problem. The cost function is completely dependent on the described task.
2. Learning Paradigm: there are three major paradigm of learning in ANN. They are

a. Supervised learning
b. Unsupervised learning
c. Reinforcement learning

1. Supervised Leaning: In this type we are given with the set of examples and the main aim for us is to find the function and map them with the examples. [3] The cost function is related to the mismatch between our mapping and the date is implicitly contains prior knowledge about the program domain.

Some of the commonly used costs are Mean squared error which minimizes the average squared error between the network's outputs. While the other minimizes the cost by gradient descent which is known multi layer perceptrons (MLP).

The task that falls under this category of learning paradigm is pattern recognition and regression. This method is capable for the sequential data input and also the quality of solution obtained so far at perfect.

2. Unsuervised learning: In this method some data is given and the cost function to the minimized that can be any function of the date and the network's output. Generally the cost function is dependent on the task and prior assumption. We arrive at the value that is equal mean of the data by minimizing the cost. This makes it much more complex in terms of cost function. [14] The task that falls under this category of paradigm are generally estimated problems, the applications that include clustering, estimation of statistical distribution, compression and filtering.

3. Reinforcement learning: In this type we are never given with the data, but eventually the data is generated agent's interaction with the environment. At each and every particular time interval agent performs certain operations and accordingly the environment generates an observation. [12] And also the cost is estimated accordingly to sum dynamics. The aim is to discover a policy for selecting actions that minimize some measures of a long term cost, Normally the environment's dynamics and the long term cost remains unknown but it can be predicted.
Eventually the environment is modeled as a MARKOV Division process (MDP). [8] These MDP consist of the following probability distributions:

1. The instantaneous cost distribution
2. The Observation Distribution
3. Transition

The main aim of MARKOV CHAIN (MC) is to discover the policy that minimizes the cost.
ANN is implemented in this type of learning Paradigm as a part of the overall algorithm. BERTSEKAS and TSITSIKINS clubbed dynamic programming with ANN to apply it in multi-dimensional non linier problems such as vehicle routing, natural resource management or medicine. As the ANN has the ability of mitigating losses of accuracy even then the discretization grid density is reduced for numerically approximating the solution of the original control problem.
The task that falls under this category of paradigm are control problems, game and other sequential decision making tasks.

Learning algorithm: It is essential to train a Neural Network by selecting one neural network methods from the set of allowed models. This is mainly done to minimize the cost criteria. [6] There are various algorithms that are available to train the Neural Networks update in which max off are viewed as straight forward application of optimization theory and statistical estimation.
Most of the algorithm that is available uses gradient descent to train the Neural Network. This is done using back propagation to compute the accurate radiant. Cost function's derivator with respect to the network parameters is found which are then changed using the parameters gradient related directions. The algorithms are usually classified in to three categories. They are:

1. Steepest discent rate and momentum, resilient
2. Quasi-Neutan
3. Levenderg-Marquardt and conjugate gradient

***Methods for Neural Networks are:***

1. Evolution methods
2. Gene expression programming
3. Stimulated Annealing
4. Expectation maximization
5. Non-parametric methods and
6. Partial swarm optimization

**Employing Artificial Neural Networks:**

The greatest advancement in the field of ANN is the ability of a system to learn which is an arbitrary function approximation mechanism which receives results from observed data. [6] However using them is no so straight forward and the relatively good understanding of the under laying theory is essential. [13]

Choice of Models: These models are dependent on the data representation and application. [15] These are capable to face the challenges in learning.

Learning algorithm: We generally have enormous number of learning algorithms. Almost all the algorithms work with the correct hyper parameters. These parameters are helpful in training a particular data set. [12] To train an unseen data we require significant amount of experimentation in choosing an algorithm.

Robustness: Once when the models costs functions and learning algorithms are selected perfectly then the outcome of ANN is extremely robust. For the past parallel implementation in hardware we just require a simple implementation and the existence of most local dependency exhibited in the treasure. [10] With an appropriate implementation of correct implementation of ANN can be used for online learning and large data set applications.

**Applications:**

The applications of ANN fall under the broad categories:

1. Function approximation or regression analysis, including time series prediction, fitness approximation and mode- ling
2. Classification, including pattern and sequence recognition, novelty deduction and sequential decision making
3. Data processing, including filtering, clustering, blind source separation and compression
4. Robotics, including directing manipulator, prosthesis.
5. Control including computer numerical control

The application area for the system includes as follows:-

1. Identifying and control
2. Quantum chemistry
3. Game Playing and decision making
4. Pattern recognition
5. Sequence recognition
6. Medical diagnosis
7. Financial application
8. Visuavalisation and
9. In e-mal span filtering

The field of ANN has helped doctors to diagnose serious cancers. An advanced machines that is used diagnose the speed of lung cancer improved the accuracy of diagnoses. This devise was named as Hybrid Lung Cancer Deduction System (HLND). The system is used to diagnose prostrate cancer. The system generates a specific model from the diagnoses made out of large number of patients with comparing to the information of one given patient. The model which is generated by the system are not out of assumption but are out of correlating the different variables. Using Neural Networks we are not able to predict colorectal cancer. [11] The outcome of these cancer patients are predicated even more accurately using ANN than the current clinical methods.

**Neural Networks and Neuro Science:** this is the field which has a concerned with theoretical analysis and computational modeling of biological Neural systems. [14] This is the field that is closely related to cognitive and behavioral modeling. The aim of the field is to create models of biological neural system in order to understand how biological systems work. [10] Many Neural scientist strive to make a link between observed biological process, biologically plausible machines for neural processing and learning for the understanding about how biological system works.

***Types of models***: Eventually there are many models that are used in the field which are defined at different levels of abstraction and modeling. [11] These models range from short term behaviors of individual neurons. Models of how the dynamics of neural circuitry arise from the communication between individual neurons and finally to models of how behavior arise from abstract neural models that represents computer sub-system, The modules includes long term and short term plasticity of neural system and there relation to learning and memory from the individual neurons to the system level.

***Networks with memory:*** The history of introducing ANN has long back traces, starting from the researches in distributed representations and self organizing maps. Deep learning is a concept in which it is useful for schematic hashing with a deep graphical model of work count vectors which are obtained from a large set of documents. [14] Generally documents are schematically located at a nearby address which is done by mapping it to the associated memory address. When the document that is to be found is a query document itself then it can be found by accessing all the addresses that deferred only a few bits from the query document.

Yet another enhancement in neural networks was the memory networks. The goal of long term memory is for prediction. [3] This was applied in the context of question answering (QA). The QA is knowledge based and the output is textual response which is an implication of long term memory. Coupling the external memory resources with deep neural networks, GOOGLE DEEP MIND came up with an system so called 'Neural Turing Machines". This machine was able to interact with by attentional process. The primary portion was capable of simple algorithms such as copying, sorting, and associative recall from input and output examples. The system is now analogous to a Turing machine but is differentially end to end, and also allow to train with gradient descent. The advancement neural Turing machine was differentiable neural computers (DNC) which was also the product of Deep Mind. The combined concept of memory networks and long-short term memory which sequential processing evolved.

**Neural Network Software:**

These software are used to stimulate, research, develop and apply ANN, biological neural networks and in some cases or wider array of adaptive system.

**Types of Artificial Neural Networks**

The types of neural networks differ from one or two layers of single direction logic to complicated multi input many dimensional feedback loop and layers. These systems use algorithms for the control and organizing the functions. Most of the system uses the parameter 'weight 'to verify the connection of the neural networks. [10] GIANNA GIAVELLI who was the first pioneer for the Neural cube style neural networks that provides dynamic space in which networks dynamically recombine information and links across billions of self adopting nodes unitizing neural Darwinism, a technique developed by GERALD EDELMIN which allows for more biologically modeled systems.

**Theoretical properties:**

Computational power: as proved by the universal approximation theorems, it is being proved that multilayer preceptron is the universal approximation. This proof was not able to convince the number of neurons acquired, the network topology the setting of the weight and the learning parameters. [7] Recurrent architecture with relational valued weight has full power of a universal Turing machines using finite number of neurons and standard liner connections has been proved by HAVA SIEGELMANN and EDUERDO D.SONTAG on their research works. [10] Super Turing power can be achieved by using irrational values for weights.

***Capacity****:* ANN has a concept called capacity which relates the amount of information that can be stored in the network and to the notation of the complexity.

***Convergence:*** It says about the number of factors and nothing more than that. [10] Firstly it consist of large number of local minima which depends on the cost function and models. Secondly the optimizations methods do not guarantee converge when far away from local minimum. Thirdly some of the tasks where the data is large number the method become impractical. Generally for practical applications, the therotical guarantees regarding convergence and they are unreliable.

***Generalization and statistics****:* The problem of over training has emerged where the goal is to create system that generalizes well in unseen examples. This arises once when the capacity of the network signiticantly exceeds the needed free parameters. [12] There are two types of schools of thought for avoiding this problem. [13] They are :

1. Cross validation and similar techniques to check for the presence of over training and optimally select hyper parameters such as the maximizing the generalizing errors.

3. Use generalization, the concept that emerges naturally in a probabilistic frame work. This can be done by selection a large prior probabability over simpler models. The goal is to minimize two quantities namely emerical risk and structural risk.

MSD can be formal statistical method to determine the confidence of the trained models in the supervised neural networks. MSE can be used as an estimate for variant when they are on a validation set. The values thus obtained can be used to calculate the confidence interval of the output probability distribution stays the same and the network is not modified the confidence analysis made of spastically valid stays the same too. To generalize the logistic function on the output layer of the neural network for categorical target variable, soft-max activation function can be assigned. The output of the function can be interpreted as posterior probabilities. [11] This is helpful in the classification as it gives a certainly measured on classification.

***Criticism:***

Training issues: one of the major criticisms in ANN mainly on the robotics is that they require an enormous and large diversity of training for real world operation. It is well known that learning machines need sufficient representative examples in order to capture the underlying structure that allows it to generalize to new cases. "Knowledge-based training of artificial neural networks for autonomous robot driving" a paper published

by Dean A. Pomerleau tells about how a neural network is trained to drive a vehicle in the multiple types of roads. [9] The large amount of researches has been made to:

- Extrapolating multiple training scenario from a single training experience and
- Presenting past training diversity so that the system does not become over-trained.

These types of issues are common in neural network and that must be decided from amongst a wide variety of responses.

***Theoretical issues:*** "Although neural nets do solve a few toy problems, their power of computation are so limited that I am suppressed anyone takes them seriously as a general problem-solving tool" was the open statement by A. K Dewdney a mathematician and computer scientist at the University of Western Ontarion and former scientist America Columnist in 1997. No neural networks update are able to solve complex problems such as n-Queens problem, travelling salesman problem or factoring large numbers etc. [12] the main drawback of ANN is that they fail to reflect how real neuron works. In most of the ANN, back- propagation plays an heart piece and is not able to perform the fundamental principle of real neuron which tells that information can only flow forward along the axon. The coding system of the information feed to a real neuron is not yet known/traced. [2] Once when the sensor activation and muscle cells pull more strongly when their associated network neurons receive action potentially more frequently, sensor neurons fire action potential are more frequent. [11] The question for how information is handled by the real neuron is known by the sensor neuron that is almost nothing but the underlying general principles unlike other simplest case of just relaying information.

The aim of the ANN is not to replicate real neural function but to use it as an inspiration for an approach to computing that is inherently parallel and which provides solutions to problems that have up until now been considered intractable. New and powerful general principle for processing information is embodied in central calm of ANN. [13] But unexpectedly the general principles are in defined. This allows simple statistical association to be described as learning or recognition. [10] As a result ANN has an impact of "something-for-nothing quality".

Hardware issue: To construct an effective ANN, considerable processing and storage resource are to be allocated. Generally a human brain has a hardware tailored to the task of processing signals through graph of Neurons. [13] It is a mile stone for a neural network designer to fill enormous number of data base rows that matches the human brain which is even more simplified from an VON NEWMANN architecture once after processing signals through graph of neurons. This system consists of large amount of computer memory and hardware disk space.

Hybrid approaches: The mixture of neural networks and symbolic approaches can better capture the mechanism of the human brain. [10] This was the research that was recorded.

***Classes and types of ANN***

1. Dynamic Neural Network
a. Feedforward Neural Network (FNN)
b. Recurrent Neural Network (RNN)
i, Hope field net work
ii.Boltzmann machine
iii,simple recurrent Net work
iv, Eco state net work
v. Long Short term memory
vi,Bi –directional RNN
vii,Hierarchical RNN
viii. Stochastic Neural Networks
c. Kohonen self organizing maps
d. Auto encoder
e. Probabilistic Neural Net work (PNN)
f. Time delay Neural Network (TDNN)
g. Regulatory feedback Network (RFNN)
2. Static Neural Network
a. Neocognitron
b. Mc Culloch – Pitts cell
c. Radial basic function network (RBF)
d. Learning vector quantization

e. Perceptron
i. Adeline model
ii. Convolutional Neural Networks (CNN)
f. Modular Neural Networks
i. Committee of machines (COM)
ii. Associative Neural Network (ASNN)
3. Memory Network
a. Google /Deep mind
b. Face book / MemNN
c. Holographic associative Memory
d. One shot associative memory
e. Neural Turing machines
f. Adoptive resonance theory
g. Hierarchical temporal memory
4. Other types of net work
a. Instantaneously trained neural networks (ITNN)
b. Spiking Neural Network (SNN)
i. Pulse coded Neural Network (PCNN)
c. Cascading Neural Networks
d. Neuro fuzzy networks (Fuzzy logic)
e. Growing Neural Gas (GNG)
f. Compositional pattern producing Networks
g. Counter propagation Networks
h. Oscillating Neural Networks
i. Hybridization Neural Networks
j. Physical Neural Networks
i. Optical Neural Networks

## II. Conclusion

This paper has totally discussed about the upcoming generation technologies that will bridge the gap totally between computing and thinking. As all science technologies have their own merits and demerits, Neural Networks are one among them. If in case all the criticism issues are solved in future the Neural Networks would be a promising technology that would reach greater heights and solves all sorts of complex problems and help human races in all means. Thus the emerging technology under construction will make a marked difference in the field of artificial intelligence.

## Reference


[1]. McCulloch, Warren; Walter Pitts (1943). "A Logical Calculus of Ideas Immanent in Nervous Activity". Bulletin of Mathematical Biophysics. **5** (4): 115–133. doi:10.1007/BF02478259.
[2]. **Jump up^** Hebb, Donald (1949). The Organization of Behavior. New York: Wiley.
[3]. **Jump up^** Farley, B.G.; W.A. Clark (1954). "Simulation of Self-Organizing Systems by Digital Computer". IRE Transactions on Information Theory. **4** (4): 76–84. doi:10.1109/TIT.1954.1057468.
[4]. **Jump up^** Rochester, N.; J.H. Holland; L.H. Habit; W.L. Duda (1956). "Tests on a cell assembly theory of the action of the brain, using a large digital computer". IRE Transactions on Information Theory. **2** (3): 80–93. doi:10.1109/TIT.1956.1056810.
[5]. **Jump up^** Rosenblatt, F. (1958). "The Perceptron: A Probabilistic Model For Information Storage And Organization In The Brain". Psychological Review. **65** (6): 386–408. doi:10.1037/h0042519. PMID 13602029.
[6]. ^ Jump up to:**a b** Werbos, P.J. (1975). Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences.
[7]. **Jump up^** Minsky, M.; S. Papert (1969). Perceptrons: An Introduction to Computational Geometry. MIT Press. ISBN 0-262-63022-2.
[8]. **Jump up^** Rumelhart, D.E; James McClelland (1986). Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Cambridge: MIT Press.
[9]. Dewdney, A.K. (1997). Yes, We Have No Neutrons: An Eye-Opening Tour through the Twists and Turns of Bad Science. New York: Wiley. p. 82.
[10]. **Jump up^** Crick, Francis (1989). "The recent excitement about neural networks". Nature. **337** (6203): 129–132. doi:10.1038/337129a0. PMID 2911347.
[11]. **Jump up^** Adrian, Edward D. (1926). "The impulses produced by sensory nerve endings". The Journal of Physiology. **61** (1): 49–72. doi:10.1113/jphysiol.1926.sp002273.
[12]. ^ Jump up to:**a b** Edwards, Chris (25 June 2015). "Growing pains for deep learning". Communications of the ACM. **58** (7): 14–16. doi:10.1145/2771283.
[13]. https://en.wikipedia.org/wiki/Artificial_neural_network