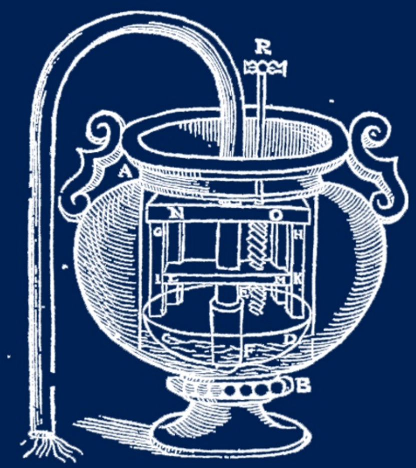
CNR-IRCrES Working Paper

Evolution of Deep Learning from Turing machine to Deep Learning next generation



4/2022

Greta Falavigna

Consiglio Nazionale delle Ricerche
IRCrES

ISSN (online) 2421-7158

Direttore	Emanuela Reale
Direzione	CNR-IRCrES *Istituto di Ricerca sulla Crescita Economica Sostenibile* Strada delle Cacce 73, 10135 Torino, Italy Tel. +39 011 3977612 / Fax +39 011 3977537 segreteria@ircres.cnr.it www.ircres.cnr.it
Sede di Roma	Via dei Taurini 19, 00185 Roma, Italy Tel. +39 06 49937809 / Fax +39 06 49937808
Sede di Milano	Via Corti 12, 20121 Milano, Italy Tel. +39 02 23699501 / Fax +39 02 23699530
Sede di Genova	Corso Ferdinando Maria Perrone 24, 16152 Genova, Italy Tel. +39 010 6598798

redazione@ircres.cnr.it
www.ircres.cnr.it/index.php/it/produzione-scientifica/pubblicazioni

CNR-IRCrES Working Paper 4/2022

Evolution of Deep Learning from Turing machine to Deep Learning next generation*

GRETA FALAVIGNA

CNR-IRCrES, National Research Council of Italy - Research Institute on Sustainable Economic Growth, Strada delle Cacce 73, 10135 Torino (TO) Italy

corresponding author: greta.falavigna@ircres.cnr.it

ABSTRACT
Despite the increasing of research papers, methodological developments, and applications of Deep Learning algorithms, a paper on the history of these models is still missing. In this study, it is provided a biography of Deep Learning, starting from its origin to this (very) moment considering the most relevant results.
The history of Deep Learning is particularly interesting; indeed, as probably never before, it was born out of the interaction of different expertise, and now we are often in touch with technologies based on these algorithms. Indeed, the first definition of neuron has been possible only for the synergy between a psychologist/neuro-anatomist, McCulloch, and a mathematician, Pitts. Together they laid the foundations for what we now call Deep Learning.
In this paper, the history with the most significant intuitions is shown, as, to our knowledge, it has never been done in previous literature. This work aims at covering this lack, presenting a chronological history of the evolution from the first neuron to today's sophisticated Evolutionary Computing, and providing the most relevant references for each addressed issue.

KEYWORDS: territorial innovation system, technology, technology innovation, R&D, startups, venture capital, industrial platforms.

DOI: 10.23760/2421-7158.2022.004

HOW TO CITE THIS WORKING PAPER
Falavigna, G. (2022). *Evolution of Deep Learning from Turing machine to Deep Learning next generation*. (CNR-IRCrES Working Paper 4/2022). Torino: Istituto di Ricerca sulla Crescita Economica Sostenibile. Available at: http://dx.doi.org/10.23760/2421-7158.2022.004

* This paper is a preliminary and preparatory outlet for the reading of the book Falavigna, G. (2022). *Deep Learning for Beginners*. Moncalieri: CNR-IRCrES (Itinerari per l'alta formazione 4). Available at http://dx.doi.org/10.23760/978-88-98193-2022-04. In details, it presents an in-depth study of the 3.1 paragraph of the cited reference.

CONTENTS

1. INTRODUCTION

Starting from the ancestors of Artificial Neural Networks, this work presents a reasoned history of Deep Learning. Although today Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), and Artificial Neural Networks (ANNs) are applied and investigated more and more, not always it is clear when they are born and who are their ancestors.

Until now, scientific literature presents interesting surveys of applications in many different fields (i.e., for imagine detection look at Zaidi et al., 2022; for finance look at Ahmed et al., 2022; for healthcare look at Egger et al., 2022; for engineering look at Watson et al., 2022 or Avci et al., 2022; for linguistic look at Pavlick, 2022), or design of new and hybrid architectures or algorithms (i.e., Albahri et al., 2022). However, as far as we know, there is not a previous paper in which milestones of DL are explained.

The main idea of this work is investigating the historical roots of DL in order to understand how and why AI arose. In general, researchers know what AI and its specifications are, but what most people ignore is that the investigated field arose thanks to a psychiatric and neurophysiologist. In addition, the ability on time in solving even more complex mathematical problems, together with the increase of computational power, allowed researchers to replicate the human brain functioning, proofing that human brain is the most perfect machine that has ever been created.

However, nowadays we are overwhelmed by concepts that are part of AI, but it is often unclear to which aspect the methods applied pertain. For this reason, before presenting the history of DL, it is necessary to clarify some definitions, in order to understand what we are talking about.

Figure 1 presents the definitions of AI, ML, and, finally, DL. AI aims to teach computers how to do what humans currently do best, therefore learning is the most important activity. The aim of AI is to build machines capable of performing characteristic tasks of human intelligence, thus simulating their cognitive capabilities. The ML offers computers the ability to learn without being explicitly programmed (Samuel, 1959) and it aims at creating AI programs capable of autonomously writing other programs to interpret data and predict results. Finally, the DL aims at improving the learning process and its goal is using ANNs to simulate the structures and functioning of the brain.

Now, it is clear that DL is a field of AI that, by using complex architectures of artificial neural networks, is able to automatically learn and reproduce complex reasoning. A recent and exhaustive definition is provided by Zhang et al. (2018) «Deep learning is a process not only to learn the relation among two or more variables, but also the knowledge that governs the relation as well as the knowledge that makes sense of the relation». In details, the DL is a special case of feature learning characterized by ANNs with two or more layers (often called multilayer or hidden layers) capable of processing information in a non-linear manner.

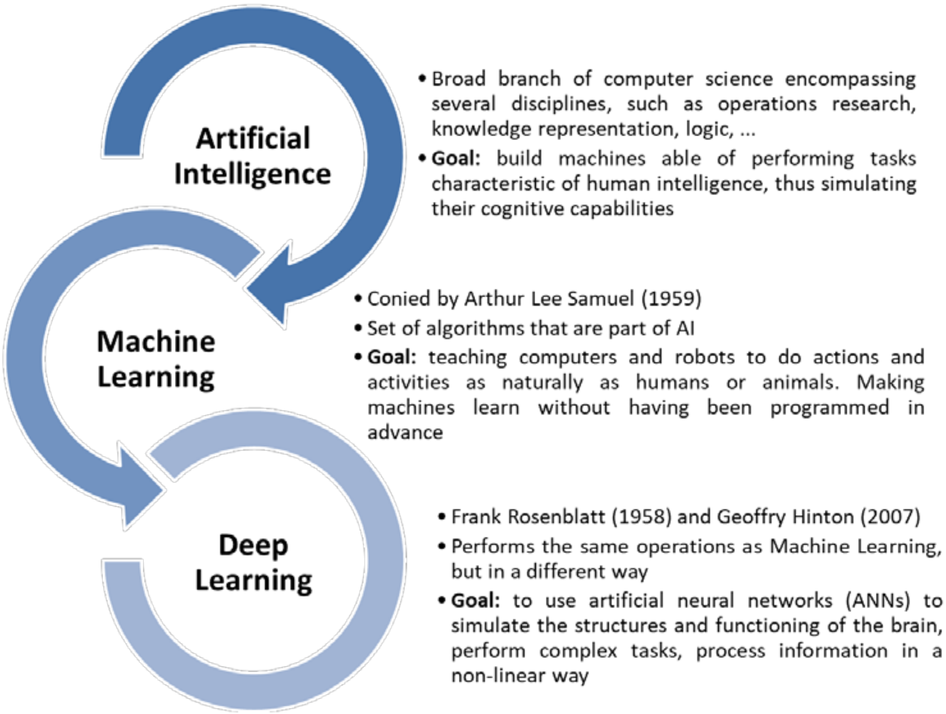


Figure 1. Definitions: Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL).

The present paper starts from the intuition of McCulloch about the function of the nervous system and continues until the last advancements in DL considering both scientific contributions and real applications.

The paper is organized as follows. Section 2 presents the years preceding the origin of the so-called Perceptron. Section 3 discusses the period between the 1950s and 1960s, while Section 4 introduces the period that is recognized as the “AI winter”. Section 5 describes the rebirth of ANNs with the definition of new algorithms, and Section 6 presents applications and new insights obtained in the last twenty years. Finally, Section 7 offers some conclusions with a particular focus on health care applications. These considerations should be of particular significance and inspiration for future works.

2. Once upon a time… a neuron…

Considering a timeline starting from the first years of the 20th century until today, we can identify that DL is a quite young field (figure 2). However, since its fundamental units are ANNs, its roots can be found before the 1950s, and even earlier, when in 1936 Alan Turing proposed a theoretical machine able to compute automatically mathematical algorithms (Turing, 1936). Since then, thanks to the increasing capability of computers and to knowledge in mathematics, researchers have made significant improvements in the study of the functioning of nervous system and of its formal description.

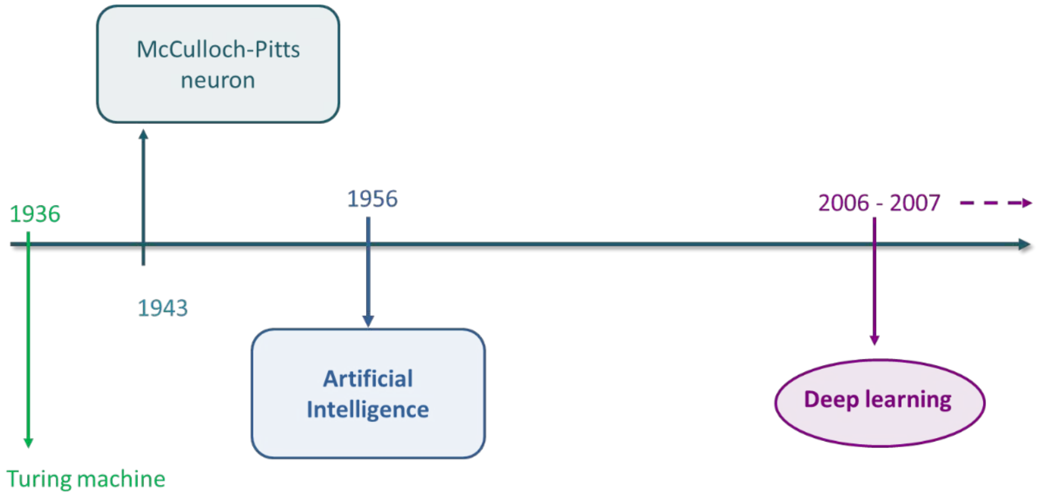


Figure 2. Timeline of the history of DL.

The first scholar to be associated with the history of neural networks is Warren Sturgis McCulloch, psychologist and neuro-anatomist, who in the 1920s began to study the functioning of the nervous system. The idea of McCulloch was to conceive a “psychon”, an elementary psychic event, representing the all or nothing impulses of neurons that, combined together, are able to produce complex events representing a proposition (McCulloch, 1961). The researcher defined some specific properties of the psychon, that can be reassumed in the following four points:

1. It had to be simple enough to have just he alternative between occurring or not;
2. It had to occur only if its cause had occurred, that is, it had to imply its temporal antecedent;
3. It had to communicate his existence to subsequent psychons;
4. Such events had to combine in such a way as to produce the equivalents of more complex propositions about their antecedents

However, the project of McCulloch found a concrete definition in 1942, when he met Walter Pitts, a mathematician who translated the idea of the psychon in algebraic terms, at the University of Chicago.

In 1943, McCulloch and Pitts published a pivotal work, in which they described the McCulloch-Pitts neuron (also called MP neuron), and they built a Turing machine with a finite number of neurons linked together. In this manner, the authors presented the neuron as the basic logical unit of the brain.

The MP neuron is represented in figure 3, where a nucleus receives signals (i.e., synapses) from inputs or from other similar neurons and produces an output on the base of the following equation and function:

$$y = \varphi(u_k) = \varphi\left(\sum_{j=1}^{m} w_{kj}x_j\right) \qquad \varphi(u_k) = \begin{cases} 1\ if\ u_k \geq 0 \\ 0\ if\ u_k < 0 \end{cases}$$

The activation function (φ) was a hardlim function, assuming only 2 values on the base of a threshold: 0 for non-active neuron and 1 for active one[2]. The synapses (w) can assume value +1 if the signal is excitatory, and -1 if the signal is inhibitory. Variables x and y are respectively

[2] An interesting and recent survey on activation function for deep learning models is provided by Dubey et al., 2022.

inputs and outputs. Even if this simple framework is the basic unit of an ANN, the MP neuron does not provide an error correction system, neither a supervised learning algorithm.

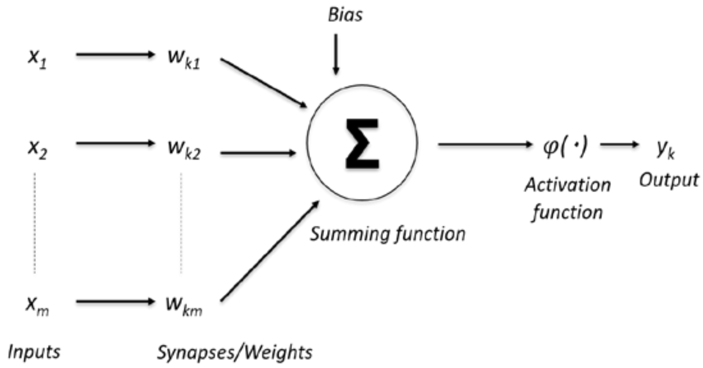


Figure 3. MP neuron (Source: adapted from Falavigna, 2022).

A few years later they developed a neural network able to recognize visual stimuli representing the same shape despite changes in orientation or size (Pitts and McCulloch, 1947). Starting from the paper of 1947, McCulloch and Pitts with a team of the Research Laboratory of Electronics of the MIT, examined the frog's visual system finding that the eye supplies the brain with information that is somehow already organized and interpreted, rather than simply transmitting an image (Lettvin et al., 1959).

In the late 1940s, the psychologist Donald Olding Hebb studied the relation between the behavior and the nervous system laying the groundwork for the unsupervised learning and the associative learning theory (Hebb, 1949). Hebb is known for the sentence "cells that fire together, wire together", meaning that "if a neuron A is close enough to a neuron B to contribute repeatedly and permanently to its excitation, then a process of growth or metabolic change takes place in both neurons such that the effectiveness of A in exciting B is heightened". This is the basic idea of the so-called Hebbian learning rule, that was a great inspiration for researchers who in the following years defined unsupervised learning algorithms Hebb's aim was to explain how the human perceptual system is able to recognize a stimulus.

3. A LUCKY DECADE FOR THE AI: 1950-1960

In this decade, experiments on MP neurons grew rapidly in number, as well as the testing of the Hebbian learning rule. It is noteworthy the work of Marvin Lee Minsky (mathematician and computer scientist) and Dean Edmonds (physicist), that in 1951 proposed the Synthetic Brain SNARC (Stochastic Neural Analog Reinforcement Calculator), through which they tested the Hebbian learning rule on 40 interconnected neurons. The goal was to build a system able of learning the path of a labyrinth. If in this case results were promising, in a laboratory of the IBM, in the same years, Nathaniel Rochester (computer scientist) and his team of researchers tried testing the Hebbian learning rule on different designs of networks, but, in this case, results were not convincing. Another interesting application was provided by Stephen Grossberg (psychologist, mathematician, and biomedical engineer), that in 1960 studied the possibility to command mechanical arms thanks to neural networks. This study represents the first formulation of the so-called "Avalanche", proposed in 1969 and 1974.

However, this decade was particularly relevant because in 1956 the Conference of Dartmouth marked formally the birth of AI. In the proposal of the Dartmouth Summer Research Project on Artificial Intelligence, John McCarthy, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon declared that "The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it" (1955). From that moment, we can consider that the term "Artificial Intelligence" has been associated with defined objectives, aimed at creating machines able of functioning as the human brain.

However, if a definition of goals of AI was identified, in this context the role of neural networks was not yet considered, and the debate on AI models became more and more lively. Indeed, we can identify two approaches: a high-level approach where intelligent programs run machines, so they are independent (Minsky is, for instance, an exponent of this current), and low-level approach where the machine itself has its own fundamental importance and intelligence strongly depends on the machine and its elementary components. This last school of thought represents the connectionist point of view, in which the leading exponent is Rosenblatt.

In these years of testing and experimentations on strengths and weaknesses of neural networks, John von Neumann (mathematician, physicist, and computer scientist) used psychologist Karl Lashley's works, who observed that mice with brain damage retain cognitive abilities. He found that different regions of the brain are equipotent from the functional point of view: any region can accomplish the given task (1950). At the same way, von Neumann, studied what happens when a part of a neural network does not work, and he suggested some solutions based on logic gate operators (1956). In addition, through the use of telegraph repeaters and valves, he was able to imitate the functioning of simple neural networks.

However, it was towards the end of the 50's and 60's that the foundations for the supervised learning of neural networks were laid, thanks to the work done by Rosenblatt. Frank Rosenblatt (psychologist) is today considered the father of DL (Tappert, 2019), because he proposed an algorithm based on error correction applied to the neuron of McCulloch-Pitts, creating the Perceptron (1958).

The Perceptron is very similar to the neuron of McCulloch-Pitts: it can assume values equal to 0 or 1, but now, a correction algorithm is defined with the aim to improve the accuracy of the neuron. This methodology is the so-called Perceptron Learning Rule (o Delta Rule, the η parameter in figure 4 is the learning rate) and allows to update synapses (i.e., weights) on the base of the ability of neuron to produce the correct output (represented by variable *t* in figure 4). The activation function and the output computation are the same of previous neuron, and in figure 4 is presented the functioning of the Perceptron Learning Rule.

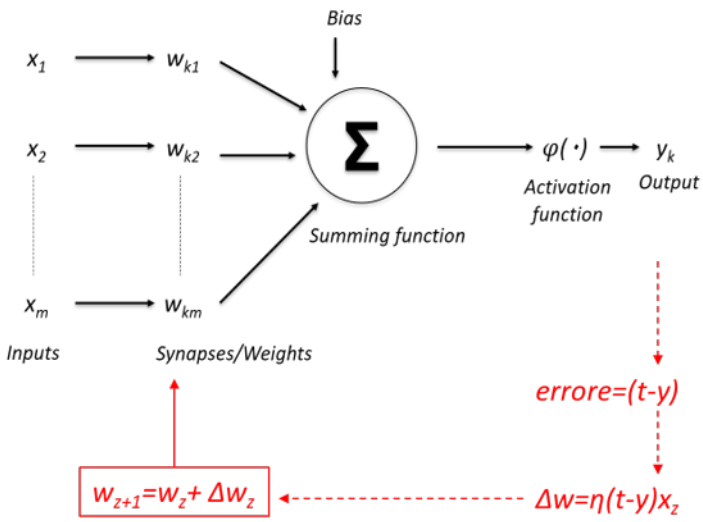


Figure 4. The Perceptron (η is the learning rate).

The first application of the Perceptron was a machine called "Mark I Perceptron". The goal was the image recognition, and it was composed by 400 photocells, randomly connected to neurons. The weights were updated while learning through electric motors.

From this moment, the supervised learning grew more and more, proposing different algorithms for the error minimization. Indeed, few years later, Bernard Widrow and Ted Hoff (electronical engineers) proposed a revised version of both neuron and learning algorithm. In details, authors proposed the ADAptive LInear NEuron (i.e., ADALINE), where the main difference with the MP neuron and Perceptron was a linear activation function, instead of a stepwise one (1960).

However, the most significant strength of this neuron is the algorithm for the supervised learning: researchers started from the perceptron learning rule and proposed the well-known “gradient descent”. Figure 5 shows the ADALINE with the gradient descent learning algorithm[3]. The function to minimize is not the simple difference between the theorical output and the empirical one, as in the case of Perceptron, but now the function to minimize is the Sum Squared Error (SSE) and the cost of loss function is defined as follows:

$$J(\boldsymbol{w}, \boldsymbol{x}) = \frac{1}{2} \sum (t - y)^2$$

In general, this approach is preferred because of differentiability of function to minimize.

In addition, the updating of weights is done on sample or sub-sample of training set (i.e., batch training), whereas the perceptron learning rule works updating weights whenever an observation is presented to the neuron (i.e., incremental mode).

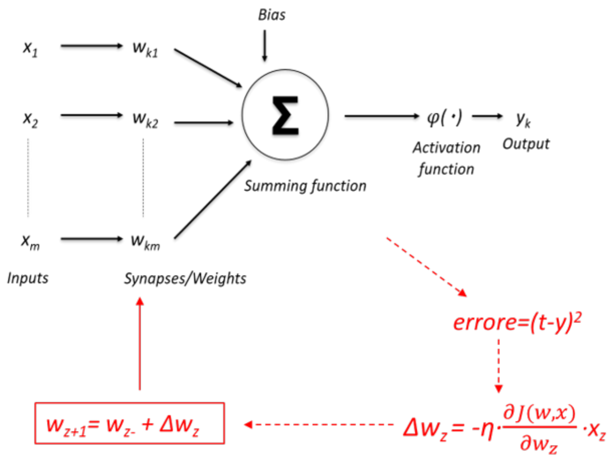


Figure 5. ADALINE and gradient descent learning algorithm (η is the learning rate).

4. AI WINTER

Unfortunately, between 1969 and 1972 ANNs were studied primarily for their weaknesses, and this period is called the “AI winter”.

Indeed, in the book “Perceptrons: An introduction to computational geometry”, Marvin Minsky (mathematicians and computer scientist) and Seymour Papert (mathematicians, computer scientist, and pedagogue) clearly identified the weaknesses of both the neuron of McCulloch-Pitts and the Perceptron (1969). Their research showed ANNs frailties from the computational point of view and, in detail, the difficulty in solving non-linearly separable functions, as well as the logical XOR gate. These theses, together with the Minsky’s belief about the superiority of the high approach to AI based on classical computer, led to the interruption of funding for research based on connectionism.

Nevertheless, in these years some steps forward have been made. James Anderson (physicist, psychologist) built a neural network model with “memory”, not very dissimilar from the ADALINE model. In his paper of 1972, the researcher supported the thesis that ANNs are able to recognize stimuli presented in the past; he proposed the idea that ANNs store information and they can recover a memory from the past.

During this period, a very relevant work was made by Kunihiko Fukushima (electronic engineer) with the definition of a multilayer neural network for pattern recognition with an unsupervised learning algorithm (1975). The “cognitron” used the strengths of the learning based on the unsupervised paradigm according to which the neural network is able to organize itself

[3] A multiple ADALINE (i.e., MADALINE) was proposed by Winter and Widrow in 1988.

according to the input data only, specializing with the aim to discriminate them on the base of the most relevant detected differences[4]. The cognitron is presented in figure 6 where four layers are not completely interconnected. Neurons are divided into inhibitory and excitatory and synapses are able to "self-organize" by selectively responding to the presentation of different patterns. The idea of Fukushima was that the first layer (i.e., U0) acted as the retina, whereas the last one (i.e., U3) acted as visual cortex.

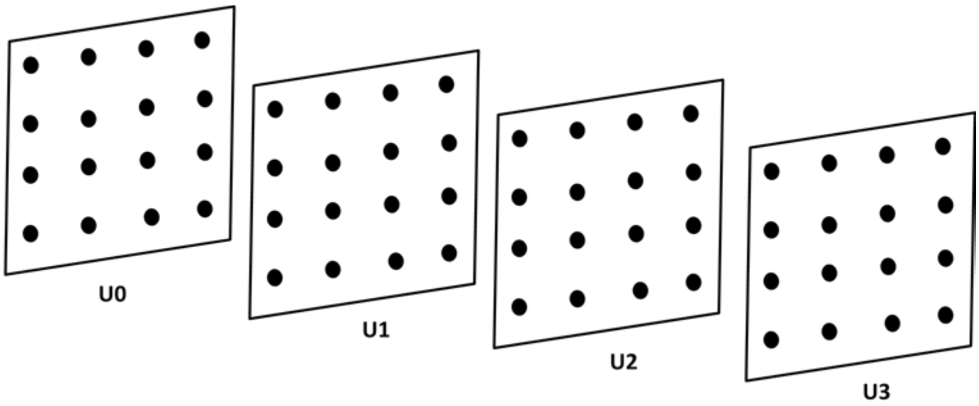


Figure 6. The cognitron

5. THE SUCCESS OF CONNECTIONISM: THE UNSUPERVISED LEARNING AND THE REINFORCEMENT LEARNING

During the 1980s and the1990s there have been many technical advancements and experiments in ANNs' architectures and especially in learning algorithms.

First of all, in 1982 the physicist John Joseph Hopfield presented the first network that behaves like an "associative memory". Starting from distorted or incomplete data, thanks to its "memory", the network is able to recover the missing information and associate it with the correct result. For instance, if images are stored in a Hopfield network, they can then be recovered by supplying similar, noisy or distorted images to the network. Hopfield networks are represented in Figure 7. They are completely inter-connected; each neuron (*h(x)*) is simultaneously input and output and it presents a binary activation state (i.e., 0 or 1). Finally, the network is associated with an "energy function" which must be minimized.

It can be shown that when this type of networks follows some properties, such as the connection of all neurons to each other or the symmetry of weights between neurons, the network, after a series of commutations, is able to end in a stable state. This means that if the network was able to store an image as a stable state, it would have been able to recover the same image even if heavily damaged.

The associative memory of Hopfield and its unsupervised mechanism inspired the following research, especially for the possibility to built more complex frameworks of networks.

Indeed, few years later, in 1984 Teuvo Kalevi Kohonen (computer scientist), proposed an unsupervised neural network model able to self-organize itself in order to form sensory maps inspired by those of somatosensory cortex. These networks consist of only two layers (input and output). The basic idea provides a competitive learning process through which neighboring neurons specialize in recognizing similar sensory stimuli. In this way, for each stimulus, the neuron with a higher level of activation is declared as the winner.

[4] An evolution of cognitron has been proposed in 1980 by Fukushima (the so-called "Neocognitron"). The neocognitron was created to recognize images even when they are transformed with roto-translations and scale variations, or when they are deformed. Finally, these networks can be considered as the ancestor of the Convolutional Neural Networks.

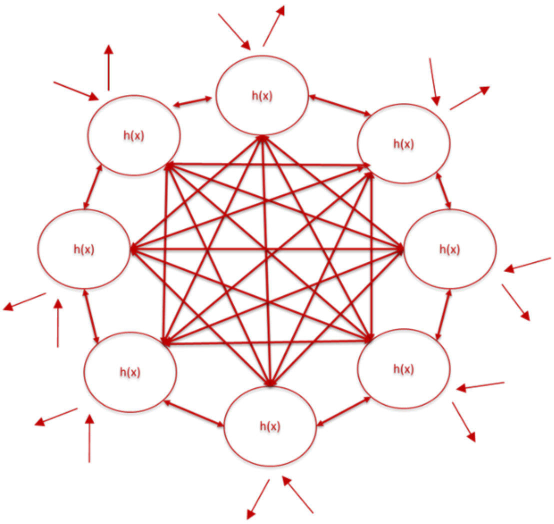
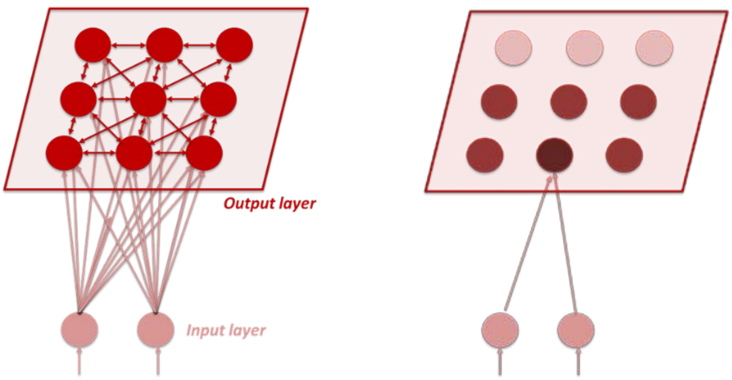


Figure 7. Hopfield network (Source: adapted from Falavigna, 2022).

The Self-Organizing Maps (SOM) are widely applied for data compression or clustering based on similarity rules. This is possible as they are able to extract the relevant information from the data without any supervision. Figure 8 (a) shows the two layers of a SOM, before the learning process. The network is completely connected and the learning process considers the connections between neurons and the influence that a neuron can have on its neighbors: neurons close to active neurons are stronger; away from neurons are weaker. As introduced before, the unsupervised learning proposed by Kohonen is a competitive learning where the neuron with the highest activation of all is the winner. Figure 8 (b) shows what happens after the learning process and a neuron is considered the winner (in the figure the darkest neuron).



(a) (b)

Figure 8. Self-Organizing Map (SOM).

While Hopfield and Kohonen studied the basis of the unsupervised learning, a team of researchers proposed a new approach to the learning process: the Reinforcement Learning (RL). Indeed, in the 1983, Andrew Barto (mathematician and computer scientist), Richard Sutton (psychologist and computer scientist), and Charles Anderson (computer scientist) proposed a new neural network model in which control actions are generated through a new learning method based on rewards and punishments. At the beginning, control actions were randomly generated by the network. Based on results from these actions, the network received a reward (positive feedback signal) or a punishment (negative feedback signal). Now, weights are updated on the base of the feedback signals received for encouraging actions with rewards and discouraging the others.

Figure 9 shows the working process of the RL algorithm.

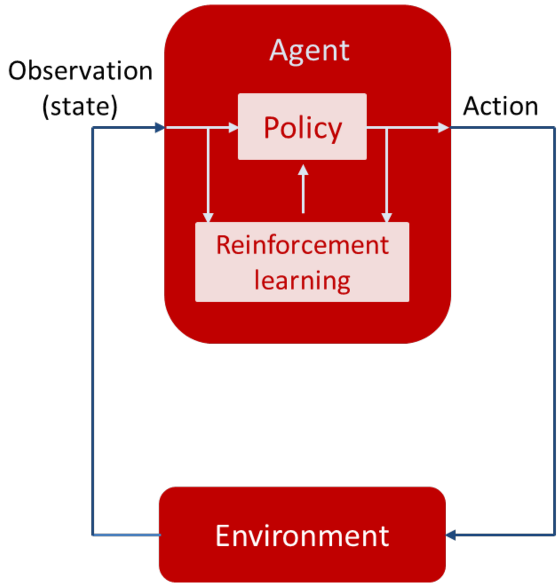


Figure 9. The RL algorithm.

RL algorithm generally follows the following seven steps:

1. Problem formulation: this phase defines the activity the agent must learn, including how the agent interacts with the environment and any primary and secondary goals the agent must achieve.
2. Environment characterization: in this phase the environment where the agent acts is created. At the same time, also the interface between the agent and the environment is decided.
3. Reward definition: specifying the reward signal used by the agent for measuring its performance compared to the goals and for evaluating how the environment has calculated the signal.
4. Agent creation: in this phase the policy representation and the learning algorithm of the agent is defined. Now, the agent is created.
5. Agent training: the representation of the policy is trained, together with the reward and the learning algorithm.
6. Agent validation: the performance of the trained agent is evaluated considering the simulated environment.
7. Policy deployment: the trained policy representation can be deployed.

Even if RL is not a particular ANN topology, it is a learning algorithm that has spread over time in an increasing number of applications based on DL, to the point that today literature presents Deep RL application (Henderson et al., 2018; Sewak, 2019). The strength of Reinforcement Learning is the possibility to scale decision-making problems that were previously not easily solved due to high dimensionality and complexity. From its first formulation, the Deep RL has been applied in many fields and the optimization algorithms have been improved in order to overcome the obstacles caused by the complexity of factors to be considered in decision-making processes (Arulkumaran et al., 2017; Wang et al., 2020).

This learning algorithm represents a very big improvement and it is applied in several field, as for example in driving automation system (Abrecht et al., 2021; Singh et al., 2022). Indeed, this method allows to modify the behavior of an agent using the system “the carrot and the stick” through rewards and punishments in order to correct the behavior of the agent.

6. THE BACKPROPAGATION FOR MULTILAYER NEURAL NETWORKS AND COMPLEX NETWORK ARCHITECTURES

In 1986 David Rumelhart (psychologist), Geoffrey Everest Hinton (computer scientist) and Ronald Williams (computer scientist) propose the supervised learning algorithm (backpropagation) applied to multilayer networks. The three researchers applied the well-known algorithm of backpropagation to a network with more hidden layers. Figure 10 presents a Feed-forward neural network where neurons of different layers are connected to each other (i.e., one neuron is connected to the neuron of the next layer), but there are no connections between neurons of the same layer nor between neurons belonging to non-adjacent layers. The neurons are very similar to the perceptron, except for the output function which is not only binary.

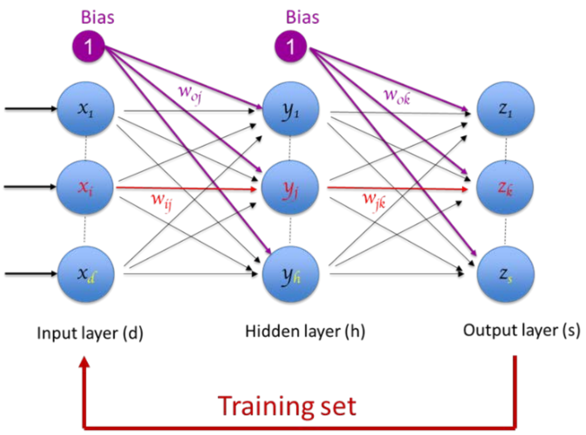


Figure 10. The backpropagation algorithm on multilayer neural network (Source: adapted from Falavigna, 2022).

This type of network is able to generalize what it has learnt in order to correctly classify new data never seen in the training phase. The main goal is to extract rules from initial sample with the aim to apply the rules to unknown data (validations and testing sets) and to verify that the neural network is able to learn from data and simulate result.

In these years, two meetings USA-Japan were held on the subject of AI, but the first official conference on neural systems was held in San Diego in 1987. 1800 people and 19 companies attended the conference. During this event was founded the International Neural Systems Society, highlighting the even more relevant role of connectionism and AI in scientific applications.

Henceforth, the increase in neural networks applications has gone hand in hand with the increase in model complexity. For instance, in 1997 the IBM's Deep Blue computer beat Russian champion Garry Kasparov at chess. More, in 1998 the National Aeronautics and Space Administration (NASA) deployed the Remote Agent, the first autonomous integrated planning system for managing spacecraft operations. In 1999: Sony launched Aibo, the AI dog, of which an upgraded version has been available in 2017.

In 1996, Jürgen Schmidhuber (computer scientist) and Josef "Sepp" Hochreiter (computer scientist) proposed a Long Short-Term Memory (LSTM) recurrent network. These typologies of networks present connections between hidden and incoming neurons, making it possible to deal with temporal sequences. For instance, an interesting survey of LSTM applications is provided by Smagulova and James (2019); instead a recent application of ANFIS and LSTM historical series of atmospheric pressure is provided by Bilgili et al. (2022).

In 1998, Yann LeCun (computer scientist), Léon Bottou (mathematician, computer scientist), Yoshua Bengio (computer scientist) and Patrick Haffner (computer scientist) defined a multilayer feedforward neural network model with a specific architecture inspired by the organization of the visual cortex (i.e., Convolutional Neural Networks, CNN). The typical topology of a CNN consists of a sequence of many hidden convolutional layers and the solution problem follows three phases.

As shown in figure 11, through filters/kernels obtained by convolutional operations, the convolutional layers extract features from the previous layer. Thanks to the linear function, only extracted features can pass to the following step that will run with a lower number of parameters. Finally, in the last layer (i.e., the fully connected) the number of neurons will be equal to the number of categories to be recognize and the process of classification will perform the final output[5].

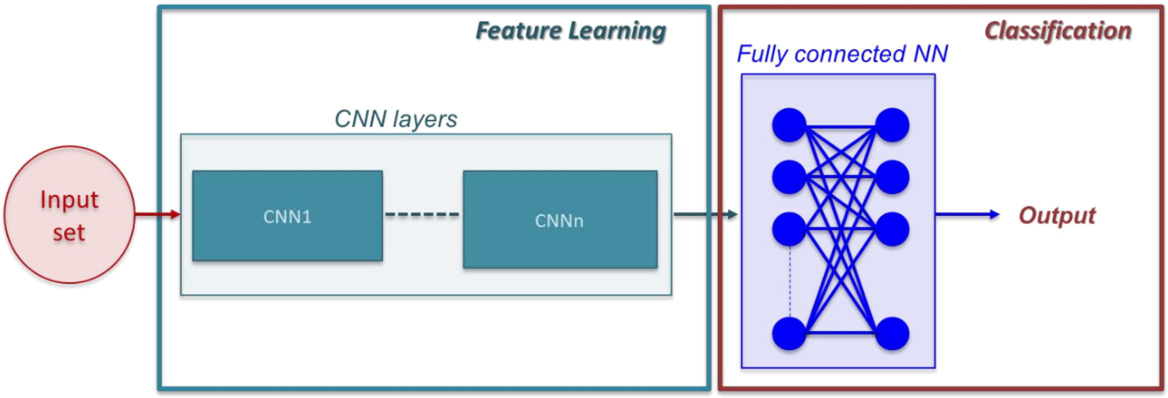


Figure 11. CNN architecture.

CNNs are even more used together with LSTM networks, jointly applied in order to minimize cost functions. In general, their jointly application is very useful in prediction and their use can be very useful in managerial process (see for instance the recent paper on mobile traffic in Milan, Eramo and Catena, 2022 or Subramanya and Riggio, 2021).

The mixture of technics represents the future of this field, together with the definition and the testing of new algorithms. However, the combination of LSTM and CNN is growing due to their characteristic to consider together the long- short-term dependencies and the accuracy in classification and prediction. Literature presents application of LSTM-CNN models in many fields, but their recent jointly use is very impressive in text mining (see for instance Luan and Lin, 2019; Ombabi et al., 2020; Zeghdaoui et al., 2021; Sirshar et al., 2022).

7. Conclusions: DL next generation

In this section, one should be able to explain what the recent uses of DL models are, but no matter how exhaustive one tries to be, one will always forget some application.

The reality is that AI systems based on DL models are becoming more and more widespread and we often do not realize it.

For instance, think about the assisted driving systems that help the driver to maintain the correct direction on the road. These are based on RL algorithms and neural networks. Or let us consider the virtual assistance technologies (e.g., Amazon's Alexa or Google Home or Apple's Siri, and so on) that we can now use simply from a phone app. Today, AI is an integral part of our lives, and scientific and applied research aims to create models that are increasingly efficient in terms of resource consumption and more effective in terms of accuracy.

Indeed, recent and future researches together with the even higher computational power aim at improving the precision of models integrating together several algorithms. History teaches that hybrid models take advantages of each technic, mitigating any existing weakness.

On this path, the most recent research suggests to take inspiration newly from biology and the evolutionist theory. Indeed, the Evolutionary Computing (EC) and Swarm Intelligence (SI), that

[5] In general, the output transfer function is a softmax ($softmax(x) = \frac{e^x}{\sum e^x}$) or is represented by a support vector machine (SVM).

include evolutionary algorithms and swarm behavior, considers optimization procedures inspired by the mechanisms of biological evolution and behaviors of swarm organisms (Zhan et al., 2022). However, nothing is new! Evolutionary algorithms have been used for neural network optimization for more than 30 years (Zhou et al., 2021), but only recently the possibility to access to big data and the improvement of computer performance, make it possible to deepen more advanced evolutionary algorithms applied to DL systems. Indeed, the accuracy of DL models' results depends strictly on parameters and, due to the large amount of available data, the necessity to find optimization algorithms is even more stringent. EC and SI are used for searching and optimizing the hyper-parameters and architectures of DL systems, providing the best solution (Darwish et al., 2020). For this reason, and due to the attention given by scholars, the next generation of DL can be identified with Evolutionary Deep Learning (Zhan et al., 2022[6]).

In AI field, theoretical research and applications go together in the same direction, and often the stimulus for methodological improvements is a practical problem. The future is geared towards the development of Information and Communication Technologies (ICT) as a driver for social, environmental and economic sustainability, and DL has a key role in developing the Internet of Things (IoT). A very relevant and recent field in which IoT solutions have a great success is the healthcare and medical science. In particular, DL techniques are integrated in IoT devices for medical diagnosis (i.e., cerebral vascular accidents; heart disease; atrial fibrillation episodes; liver disease; treatment for lung cancer, and so on) using the strengths of CNNs and LSTM in imaging detection (Tuli et al., 2020; Rebouças Filho, 2017; Masoumi et al., 2012; Faust et al., 2018; Schirrmeister et al., 2017; Sung et al., 2021). Home-based and personal healthcare applications are built with DL algorithms, as suggested by Fonseca et al. (2018) that propose intelligent living spaces for chronic patients with multimorbidity. Another interesting application is provided by Liu et al. (2019) presenting a platform based of IoT and DL for detection and classification of teeth disorders. A very promising application concerns the physiological monitoring based on Body Sensor Network (BSN) with the aim to control and predict the health status, mostly of elderly people (Sharma et al., 2018; Zhang et al., 2015; Huang et al., 2018; Braunstein, 2018). In this stream, we can also think to common applications as, for instance, smart technology for diet predictions. We already use smartwatches that monitor our physical activity by showing the number of steps, but also the calories consumed. Of course, different types of these devices exist, with different sensitivities depending on the software used. The basic idea, however, is to use sensors to monitor people's activity and suggest a customized diet, useful not only in cases of obesity, but also to prevent diseases related to incorrect eating habits (Wei and Cheok, 2012; Islam et al., 2016). All these applications rely on the great and acknowledged generalizing capacity of ANNs, as well as the classification and prediction skills of the most refined and recent DL algorithms. However, the literature suggests that these models suffer until now of overspecialization on patterns and sample, but the increasing availability of big data in the next future probably could overcome this weakness. A recent and exhaustive review of literature about IoT and DL applied in the health care field is provided by Bolhasani et al. (2021).

Undoubtedly, we must be ready for a rapidly increasing of IoT even in the simplest situations of our lives. Indeed, as suggested in a very recent work of Bhattacharya et al. (2022), in the not too distant future our cities will increasingly resemble smart cities in which AI and DL models will be able to help people or replace them in certain tasks.

Finally, if, from the one hand the presented technologies can surely improve the life quality, especially considering that they represent only a small part of the applications already existing and under study, from the other hand, the pressure towards automation and the replacement of human work by machines is also likely to change the quality of social relations (think of the growing use of social networks such as Facebook, Instagram, Twitter, and so on...).

What can be defined as the optimal balance between "computer thinking" and "human thinking" will always remain an undefined, and it is mainly in this direction that future researches

[6] The authors (Zhan et al., 2022) provides a very interesting and useful survey of Evolutionary Deep Learning models and, therefore, they define a systematic taxonomy of technics.

should go: reaching a break-even point that allows people employing the strengths of technology, without risking a de-humanization of social relations.

8. References

Abrecht, S., Gauerhof, L., Gladisch, C., Groh, K., Heinzemann, C., & Woehrle, M. (2021). Testing deep learning-based visual perception for automated driving. *ACM Transactions on Cyber-Physical Systems (TCPS)*, *5*(4), pp. 1-28.

Ahmed, S., Alshater, M. M., El Ammari, A., & Hammami, H. (2022). Artificial intelligence and machine learning in finance: A bibliometric review. *Research in International Business and Finance*, *61*, 101646.

Albahri, A.S., Alnoor, A., Zaidan, A. A., Albahri, O. S., Hameed, H., Zaidan, B.B., ... & Yass, A.A. (2022). Hybrid artificial neural network and structural equation modelling techniques: a survey. *Complex & Intelligent Systems*, *8*(2), pp. 1781-1801.

Anderson, J. A. (1972). A simple neural network generating an interactive memory. *Mathematical biosciences*, *14*(3-4), pp. 197-220.

Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, *34*(6), pp. 26-38.

Avci, O., Abdeljaber, O., & Kiranyaz, S. (2022). An overview of deep learning methods used in vibration-based damage detection in civil engineering. In Grimmelsman, K. (eds). *Dynamics of Civil Structures, Volume 2* (pp. 93-98). Conference Proceedings of the Society for Experimental Mechanics Series. Cham: Springer. https://doi.org/10.1007/978-3-030-77143-0_10

Barto, A.G., Sutton, R.S., & Anderson, C.W. (1983). Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE transactions on systems, man, and cybernetics SMC*, *13*(5), pp. 834-846.

Bhattacharya, S., Somayaji, S.R.K., Gadekallu, T.R., Alazab, M., & Maddikunta, P.K.R. (2022). A review on deep learning for future smart cities. *Internet Technology Letters*, *5*(1), e187.

Bilgili, M., Ilhan, A., & Ünal, Ş. (2022). Time-series prediction of hourly atmospheric pressure using ANFIS and LSTM approaches. *Neural Computing and Applications*, 34, pp. 15633-15648.

Bolhasani, H., Mohseni, M., & Rahmani, A. M. (2021). Deep learning applications for IoT in health care: A systematic review. *Informatics in Medicine Unlocked*, 23, 100550.

Braunstein, M. L. (2018). Healthcare in the age of interoperability: the promise of fast healthcare interoperability resources. *IEEE pulse*, 9(6), pp. 24-27.

Darwish, A., Hassanien, A. E., & Das, S. (2020). A survey of swarm and evolutionary computing approaches for deep learning. *Artificial intelligence review*, *53*(3), pp. 1767-1812.

Dubey, S. R., Singh, S. K., & Chaudhuri, B. B. (2022). Activation Functions in Deep Learning: A comprehensive Survey and Benchmark. *Neurocomputing*, 503, pp. 92-108.

Egger, J., Gsaxner, C., Pepe, A., Pomykala, K. L., Jonske, F., Kurz, M., ... & Kleesiek, J. (2022). *Computer Methods and Programs in Biomedicine*, 221, 106874.

Eramo, V., & Catena, T. (2022). Application of an Innovative Convolutional/LSTM Neural Network for Computing Resource Allocation in NFV Network Architectures. *IEEE Transactions on Network and Service Management*, *19*(3), pp. 2929-2943.

Falavigna, G. (2022). Deep Learning for Beginners. Moncalieri: CNR-IRCrES (Itinerari per l'alta formazione 4). Available at: https://www.ircres.cnr.it/images/iaf/IAF_04_2022.pdf

Faust, O., Shenfield, A., Kareem, M., San, T. R., Fujita, H., & Acharya, U. R. (2018). Automated detection of atrial fibrillation using long short-term memory network with RR interval signals. *Computers in biology and medicine*, 102, pp. 327-335.

Fonseca C, Mendes D, Lopes M, Romão A, Parreira P. (2017). Deep Learning and IoT to Assist Multimorbidity Home Based Healthcare. *Journal of Health & Medical Informatics*, *8*(3), 273.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36, pp. 193-202.

Fukushima, K. (1975). Cognitron: a self-organizing multi-layered neural network. *Biological Cybernetics*, 20, pp. 121-136.

Grossberg, S. (1969). Some networks that can learn, remember, and reproduce any number of complicated space-time patterns. *Journal of Mathematics and Mechanics, 19*(1), pp. 53-91.
Grossberg, S. (1974). Classical and Instrumental Learning by Neural Networks. In: Studies of Mind and Brain (pp. 51-141).. Dordrecht: Springer (Boston Studies in the Philosophy of Science 70).
Hebb, D.O. (1949). *The Organization of Behavior*. New York: Wiley & Sons.
Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D., & Meger, D. (2018). Deep reinforcement learning that matters. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1).
Hochreiter, S., & Schmidhuber, J. (1996). LSTM can solve hard long time lag problems. *Advances in neural information processing systems*, 9.
Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, *79*(8), pp. 2554-2558.
Huang, H., Gong, T., Ye, N., Wang, R., & Dou, Y. (2017). Private and secured medical data transmission and analysis for wireless sensing healthcare system. *IEEE Transactions on Industrial Informatics*, *13*(3), pp. 1227-1237.
Korhonen, T. (1984). Self-organization and associative memory. Heidelberg: Springer-Verlag (series in Information Sciences, 8). [updated version: Kohonen, T. (2012)].
Islam, S. R., Uddin, M.N., & Kwak, K. S. (2016). The IoT: Exciting possibilities for bettering lives: Special application scenarios. *IEEE Consumer Electronics Magazine*, *5*(2), pp. 49-57.
Lashley, K.S. (1950). In search of the engram. In Society for Experimental Biology, Physiological mechanisms in animal behavior (pp. 454-482). Society's Symposium IV. Washington: Academic Press.
LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), pp. 2278-2324.
Lettvin, J.Y., Maturana, H.R., McCulloch, W.S., & Pitts, W.H. (1959). What the frog's eye tells the frog's brain. *Proceedings of the Institute of Radio Engineers*, 47(11), pp. 1940-1951.
Liu, L., Xu, J., Huan, Y., Zou, Z., Yeh, S. C., & Zheng, L. R. (2019). A smart dental health-IoT platform based on intelligent hardware, deep learning, and mobile terminal. *IEEE journal of biomedical and health informatics*, 24(3), pp. 898-906.
Luan, Y., & Lin, S. (2019). Research on text classification based on CNN and LSTM. In 2019 IEEE international conference on artificial intelligence and computer applications, pp. 352-355.
Masoumi, H., Behrad, A., Pourmina, M. A., & Roosta, A. (2012). Automatic liver segmentation in MRI images using an iterative watershed algorithm and artificial neural network. *Biomedical signal processing and control*, *7*(5), pp. 429-437.
McCarthy, J., Minsky, M.L., Rochester, N., & Shannon, C.E. (2018). A proposal for the Dartmouth summer research project on artificial intelligence (1955). Reprinted online at http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html
McCulloch, W.S. (1961). What is a number, that a man may know it, and a man, that he may know a number. *General Semantics Bulletin*, *26*(27), pp. 7-18.
McCulloch, W.S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, *5*(4), pp. 115-133.
Minsky, M., & Papert, S. (1969). *Perceptrons: An introduction to computational geometry*. Cambridge: Mass, HIT, pp. 479-480.
Ombabi, A. H., Ouarda, W., & Alimi, A. M. (2020). Deep learning CNN-LSTM framework for Arabic sentiment analysis using textual information shared in social networks. *Social Network Analysis and Mining*, 10(1), pp. 1-13.
Pavlick, E. (2022). Semantic structure in deep learning. *Annual Review of Linguistics*, 8, pp. 447-471.
Pitts, W., & McCulloch, W. S. (1947). How we know universals the perception of auditory and visual forms. *The Bulletin of mathematical biophysics*, *9*(3), pp. 127-147.
Rebouças Filho, P. P., Sarmento, R. M., Holanda, G. B., & de Alencar Lima, D. (2017). New approach to detect and classify stroke in skull CT images via analysis of brain tissue densities. *Computer methods and programs in biomedicine*, 148, pp. 27-43.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, *65*(6), pp. 386.
Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), pp. 533-536.
Samuel, A. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, *3*(3), pp. 210-229.
Schirrmeister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggensperger, K., Tangermann, M., ... & Ball, T. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. *Human brain mapping*, *38*(11), pp. 5391-5420.
Sewak, M. (2019). *Deep reinforcement learning*. Singapore: Springer Singapore.
Sharma, S., Chen, K., & Sheth, A. (2018). Toward practical privacy-preserving analytics for IoT and cloud-based healthcare systems. *IEEE Internet Computing*, *22*(2), pp. 42-51.
Singh, V., Chen, S. S., Singhania, M., Nanavati, B., & Gupta, A. (2022). How are reinforcement learning and deep learning algorithms used for big data based decision making in financial industries. A review and research agenda. *International Journal of Information Management Data Insights*, *2*(2), 100094.
Sirshar, M., Paracha, M.F.K., Akram, M.U., Alghamdi, N.S., Zaidi, S.Z.Y., & Fatima, T. (2022). Attention based automated radiology report generation using CNN and LSTM. *PloS one*, *17*(1), e0262209.
Smagulova, K., & James, A.P. (2019). A survey on LSTM memristive neural network architectures and applications. *The European Physical Journal Special Topics*, *228*(10), pp. 2313-2324.
Subramanya, T., & Riggio, R. (2021). Centralized and federated learning for predictive VNF autoscaling in multi-domain 5G networks and beyond. *IEEE Transactions on Network and Service Management*, *18*(1), pp. 63-78.
Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, *71*(3), pp. 209-249.
Tappert, C.C. (2019). Who is the father of deep learning?. *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 343-348.
Tuli, S., Basumatary, N., Gill, S.S., Kahani, M., Arya, R. C., Wander, G.S., & Buyya, R. (2020). HealthFog: An ensemble deep learning based Smart Healthcare System for Automatic Diagnosis of Heart Diseases in integrated IoT and fog computing environments. *Future Generation Computer Systems*, 104, pp. 187-200.
Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Journal of Math*, 58, pp. 345-363.
Von Neumann, J. (1956). Probabilistic logics and the synthesis of reliable organisms from unreliable components. *Automata studies*, *34*(34), pp. 43-98.
Wang, H.N., Liu, N., Zhang, Y.Y., Feng, D.W., Huang, F., Li, D.S., & Zhang, Y.M. (2020). Deep reinforcement learning: a survey. *Frontiers of Information Technology & Electronic Engineering*, *21*(12), pp. 1726-1744.
Watson, C., Cooper, N., Palacio, D.N., Moran, K., & Poshyvanyk, D. (2022). A Systematic Literature Review on the Use of Deep Learning in Software Engineering Research. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, *31*(2), pp. 1-58.
Wei, J., & Cheok, A.D. (2012). Foodie: Play with your food promote interaction and fun with edible interface. *IEEE Transactions on Consumer Electronics*, *58*(2), pp. 178-183.
Widrow, B., & Hoff, M.E. (1960). *Adaptive switching circuits*. Technical Report No. 1553-1 (Arlington, VA: Defense Technical Information Center – Stanford Electronic Laboratories.
Winter, C.R., & Widrow, B. (1988). MADALINE RULE II: A training algorithm for neural networks. In *Second Annual International Conference on Neural Networks*.
Zaidi, S.S.A., Ansari, M.S., Aslam, A., Kanwal, N., Asghar, M., & Lee, B. (2022). A survey of modern deep learning based object detection models. *Digital Signal Processing*, 126, 103514.

Zeghdaoui, M.W., Boussaid, O., Bentayeb, F., & Joly, F. (2021). Medical-Based Text Classification Using FastText Features and CNN-LSTM Model. In *International Conference on Database and Expert Systems Applications*. Cham: Springer, pp. 155-167.
Zhan, Z.H., Li, J.Y., & Zhang, J. (2022). Evolutionary deep learning: A survey. *Neurocomputing*, 483, pp. 42-58.
Zhang, W.J., Yang, G., Lin, Y., Ji, C., & Gupta, M.M. (2018). On definition of deep learning. In *2018 World automation congress (WAC)*, pp. 1-5.
Zhang, Y., Qiu, M., Tsai, C. W., Hassan, M. M., & Alamri, A. (2015). Health-CPS: Healthcare cyber-physical system assisted by cloud and big data. *IEEE Systems Journal*, *11*(1), pp. 88-95.
Zhou, X., Qin, A. K., Sun, Y., & Tan, K. C. (2021, June). A survey of advances in evolutionary neural architecture search. In *2021 IEEE Congress on Evolutionary Computation (CEC)*, pp. 950-957.

ISSN (online) 2421-7158

CNR-IRCrES Working Papers

2022

N.3/2022 A Simulation Model of Technology Innovation of a Territory. Angelo Bonomi.

N. 2/2022 Technology and Environmental Policies. Angelo Bonomi.

N. 1/2022 Le donne marittime: fra stereotipi di genere, discriminazioni e scarse opportunità occupazionali. Barbara Bonciani, Silvia Peveri.

2021

N. 9/2021 Management of open access research infrastructures in large EU projects: the "CultureLabs" case. Andrea Orazio Spinello, Danilo Giglitto, Eleanor Lockley.

N. 8/2021 Francia-Italia: parole in campo. Intorno alla narrazione del Campionato del Mondo di calcio donne. Antonella Emina.

N. 7/2021 Covid-19 e rischio di insolvenza: il punto di vista del mercato azionario. Franco Varetto.

N. 6/2021 Institutional efficiency and budget constraints: a Directional Distance Function approach to lead a key policy reform. Greta Falavigna, Roberto Ippoliti.

N. 5/2021 Different waves and different policy interventions in 2020 Covid-19 in Italy: did they bring different results?. Mario Nosvelli.

N. 4/2001 On Search of a General Model of Technology Innovation. Angelo Bonomi.

N. 3/2021 Design and implementation of a web survey on the effects of evaluation on academic research. Andrea Orazio Spinello, Emanuela Reale, Antonio Zinilli.

N. 2/2021 An online survey on the effects of agile working in Italian Public Research Organisations. Serena Fabrizio, Valentina Lamonica, Andrea Orazio Spinello.

N. 1/2021 Technology Transfer Activities in Universities and Public Research Organizations: A Literature Overview. Ugo Finardi, Rolfo Secondo, Isabella Bianco.

2020

N. 12/2020 Unexpected loss multiperiodale e pricing del rischio di credito. Franco Varetto.

N. 11/2020 La ricerca in Nanotecnologie e Nanoscienze in Italia: spesa del settore pubblico e aree tematiche prevalenti. Ugo Finardi, Andrea Orazio Spinello.

N. 10/2020 Persistent fast growth and profitability. Lucio Morettini, Bianca Potì, Roberto Gabriele.

N. 9/2020 Binomio *Burnout* e *Mindfulness* nelle organizzazioni. Alcuni studi e scenari di applicazione. Oriana Ippoliti, Riccardo Briotti, Bianca Crocamo, Antonio Minopoli.

N. 8/2020 Innovation and communication of companies on Twitter before and during COVID-19 crisis. José N. Franco-Riquelme, Antonio Zinilli, Joaquín B. Ordieres-Meré and Emanuela Reale.

N. 7/2020 The proposal of a new hybrid methodology for the impact assessment of energy efficiency interventions. An exploratory study. Monica Cariola, Greta Falavigna.

N. 6/2020 The technology innovative system of the Silicon Valley. Angelo Bonomi.

N. 5/2020 Storia dell'industria delle macchine utensili in Piemonte dalle origini alla seconda guerra mondiale. Secondo Rolfo.

N. 4/2020 Blockchain e Internet of Things per la logistica Un caso di collaborazione tra ricerca e impresa. Edoardo Lorenzetti, Lucio Morettini, Franco Mazzenga, Alessandro Vizzarri, Romeo Giuliano, Paolo Peruzzi, Cristiano Di Giovanni.

N. 3/2020 L'impatto economico e fiscale di un evento culturale: misure e scala territoriale. Giovanna Segre, Andrea Morelli.

N. 2/2020 Mapping the tangible and intangible elements of the historical buildings and spaces. Edoardo Lorenzetti, Nicola Maiellaro.

N. 1/2020 Il lavoro agile negli enti pubblici di ricerca. Emanuela Reale, Serena Fabrizio, Andrea Orazio Spinello.

2019

N. 6/2019 Women's candidatures in local elections: does the context matter? Empirical evidence from Italian municipalities. Igor Benati, Greta Falavigna, Lisa Sella.

N. 5/2019 Research activities in Nanotechnologies and Nanosciences: an analysis of Piedmont's nanotech research system. Ugo Finardi.

N. 4/2019 Xylella fastidiosa: patogenesi, danni economici e lotta al disseccamento rapido dell'olivo. Maurizio Conti.

N. 3/2019 Flussi di traffico attraverso il tunnel automobilistico del Frejus: un semplice esercizio di *forecasting* e alcune considerazioni a margine. Ugo Finardi.

Numeri precedenti/Previous issues

Despite the increasing of research papers, methodological developments, and applications of Deep Learning algorithms, a paper on the history of these models is still missing. In this study, it is provided a biography of Deep Learning, starting from its origin to this (very) moment considering the most relevant results.
The history of Deep Learning is particularly interesting; indeed, as probably never before, it was born out of the interaction of different expertise, and now we are often in touch with technologies based on these algorithms. Indeed, the first definition of neuron has been possible only for the synergy between a psychologist/neuro-anatomist, McCulloch, and a mathematician, Pitts. Together they laid the foundations for what we now call Deep Learning.
In this paper, the history with the most significant intuitions is shown, as, to our knowledge, it has never been done in previous literature. This work aims at covering this lack, presenting a chronological history of the evolution from the first neuron to today's sophisticated Evolutionary Computing, and providing the most relevant references for each addressed issue.

CNR - Consiglio Nazionale delle Ricerche
IRCrES - Istituto di Ricerca sulla Crescita Economica Sostenibile