Who is the Father of Deep Learning?

Charles C. Tappert

Seidenberg School of CSIS, Pace University, Pleasantville, New York 10570
ctappert@pace.edu

Abstract— This paper evaluates candidates for the father of deep learning. We conclude that Frank Rosenblatt developed and explored all the basic ingredients of the deep learning systems of today, and that he should be recognized as a Father of Deep Learning, perhaps together with Hinton, LeCun and Bengio who have just received the Turing Award as the fathers of the deep learning revolution.

Keywords—artificial intelligence, machine learning, deep learning, neural networks, multi-layer perceptrons

I. Introduction

Deep learning is important because it is causing a revolution in artificial intelligence, electrifying the computing industry, and basically Transforming corporate America [12]. It has done this over the last 5-10 years, during which time we have experienced quantum leaps in the quality of many everyday technologies. Deep learning has made major advances in image recognition. It has made possible the search for and automatic organizing of collections of photos by Apple, Amazon, Microsoft, and Facebook. It has made speech technologies work much better as shown by advances in the speech interfaces of Apple's Siri, Amazon's Alexa, Microsoft's Cortana, and Chinese Baidu. It has facilitated the Translation of spoken sentences as exemplified by Google Translate. It has also improved medical applications, robotics, autonomous drones, self-driving cars, etc.

The purpose of this paper is to review the candidates for "the father of deep learning." It is important to give credit to the inventor or creator of a technology. In the area of computing, for example, Charles Babbage, Alan Turing, John von Neuman, John Mauchley, and Presper Eckert are considered fathers of the computer; Ada Lovelace the mother of computer programming; Nathaniel Rochester the father of the assembler; John Backus the father of the compiler; and Tim Berners-Lee the father of the World Wide Web [19]. The key creators of a technology are often not recognized as fathers or mothers of the technology while they are still living.

In the following sections, this paper discusses deep learning and its related technologies, the history of deep learning and its current impact, the major types of deep learning systems, an overview of the candidates that might be considered a father of deep learning, a discussion of the major candidates and why they might be considered the fathers of deep learning, and finally some conclusions.

II. Deep Learning

A. Deep learning and related technologies

The technologies of artificial intelligence, machine learning, and deep learning are highly overlapping (Fig. 1). Artificial Intelligence (AI) refers to an artificial creation of human-like intelligence. So far AIs have been in specific narrow areas:

- In 1997 IBM's Deep Blue beat the world's best human chess player
- In 2011 IBM's Watson beat the world's best human jeopardy players
- In 2017 Google's AlphaGo beat the world's best Go player
- The Google search engine quickly finds information
- Apple's Siri and Amazon's Alexa voice services

AI has been around for many years and there is currently a renewed focus on Artificial General Intelligence (AGI), which are AIs that can perform any intellectual task that a human can. Machine Learning (ML) is a sub-area under AI and Deep Learning (DL) is a sub-area under ML. With advances in computing power, applications in these areas can now effectively utilize large quantities of data (Big Data), thus enhancing the capabilities of data mining and data science that use ML algorithms.

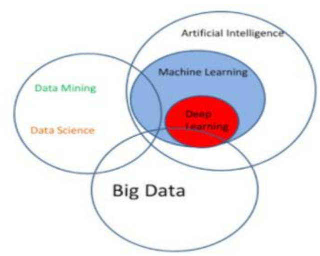

Fig. 1. AI, ML, DL, Data Mining, Data Science, and Big Data [2].

Deep Neural Networks (DNN) are neural networks having multiple layers between the input and output layers (Fig. 2).

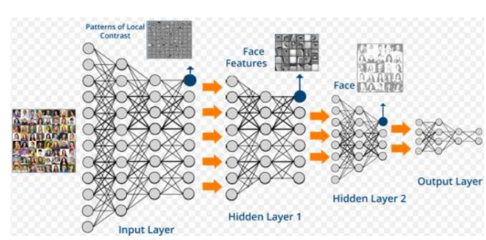

Fig. 2. A DNN for face recognition [6].

978-1-7281-5584-5/19/$31.00 ©2019 IEEE
DOI 10.1109/CSCI49370.2019.00067

It is this newly developed technology of DL that is causing real-world impact by winning the recent contests and outperforming other ML technologies in most applications, such as in the areas of machine vision and speech recognition (Siri and Alexa). For example, DL systems have won the major visual image classification contests in recent years. A dramatic moment in the meteoric rise of DL came when a convolutional network won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) for the first time, and this challenge is now consistently won by deep networks [9], see Fig. 3.

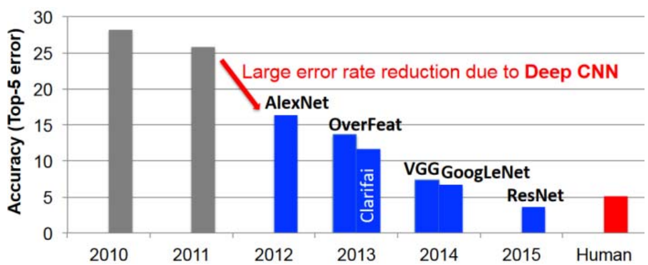

Fig. 3. DL networks caused significant error-rate drops in 2012 [9].

B. History of Deep Learning

Goodfellow, Bengio & Courville [5] describe a Three Wave Development of Deep Learning. The first wave, 1940s-1960s, consisted of early neural networks, and was termed the Cybernetics wave. This wave consisted primarily of Rosenblatt's perceptron which was developed from Hebb's synaptic strengthening ideas and the McCulloch-Pitts Neuron. The key idea was variations of stochastic gradient descent. This wave was killed by a book entitled Perceptrons [10], which basically stopped all research in neural networks for 15 years and this period is referred to as an "AI Winter." The second wave, 1980s-1990s, began with the discovery of backpropagation by Rumelhart and others, and was termed the Connectionism wave. Backpropagation made possible the training of multi-layer neural network, a key element missing from the Rosenblatt era. The third wave, 2006-present, likely started with Hinton's deep belief network, and is termed Deep Learning. The key idea of deep learning is the hierarchy of many layers in the neural network.

C. Major types of deep learning systems

The major types of DL networks are described in the popular textbook [5].

- Feed forward networks, often referred to as multilayer perceptrons, the most popular being the convolutional networks inspired by the organization of the animal visual cortex, see Fig. 2.
- Recurrent Neural Networks, a class of artificial neural network where connections between units form a directed cycle (feedback), primarily used for sequential data.

III. Overview of Candidates

There are a number of candidates for the father of deep learning and many of the candidates are still working in the area. They are reviewed in order of their birth.

A. Warren McCulloch (1898-1969)

McCulloch, a neurophysiologist, was known for his work on the foundation for certain brain theories and his contribution to the area of cybernetics. McCulloch, together with Walter Pitts, created computational neural models. These models focused both on the biological processes in the brain and on the application of neural networks to artificial intelligence [1].

B. W. Ross Ashby (1903-1972)

Ross Ashby was an English psychiatrist and a pioneer in cybernetics. His two books, *Design for a Brain* and *An Introduction to Cybernetics*, were landmark works that introduced logical thinking into the new discipline of cybernetics. His work was widely influential within cybernetics, systems theory and, more recently, complex systems [3].

C. Donald Hebb (1904-1985)

Donald Hebb was a Canadian psychologist who was influential in the area of neuropsychology, where he sought to understand how the function of neurons contributed to psychological processes such as learning. He is best known for his theory of Hebbian learning, which he introduced in his classic 1949 work *The Organization of Behavior*. He has been described as the father of neuropsychology and neural networks. His views on learning described behavior and thought in terms of brain function, explaining cognitive processes in terms of connections between neuron assemblies [4].

D. Frank Rosenblatt (1928-1971)

Rosenblatt was a psychologist best known for the perceptron, a neural network constructed in accordance with biological principles and showed an ability to learn. While at Cornell Aeronautical Laboratory in the late 1950s, Rosenblatt's perceptrons were initially simulated on an IBM 704 computer and this led to the construction of an electronic machine, The Mark I Perceptron, which currently resides in the Smithsonian Institution. After moving to Cornell University in 1960, he developed and extended this approach in numerous papers and a book (Rosenblatt, 1962) which he used as a textbook in his courses. He received international recognition for the perceptron, and The New York Times billed it as a revolution, causing many researchers to have big expectations on what perceptrons could do. Unfortunately, the book Perceptrons (Minsky & Papert, 1969) appeared in 1969 with a mathematical proof that two-layer feed-forward linear perceptrons with one trainable set of weights cannot model non-linear transformations. Although this result was trivial, the book had a pronounced effect on research funding, and led to a 15-year hiatus of neural network research. Well after his death, research on neural networks returned to the mainstream in the 1980s, and new researchers studied Rosenblatt's work. Most of these new researchers interpreted as a contradiction the hypotheses presented in the book Perceptrons, and a confirmation of Rosenblatt's expectations [11].

E. *David Rumelhart (1942-2011)*

Rumelhart was a psychologist who made many contributions to the formal analysis of human cognition, working primarily within the frameworks of mathematical psychology, symbolic artificial intelligence, and parallel distributed processing. He was the first author of the highly cited paper (Rumelhart, Hinton & Williams, 1986) that applied the back-propagation algorithm (also known as the reverse mode of automatic differentiation published by Seppo Linnainmaa in 1970) to multi-layer neural networks. This work showed through experiments that such networks can learn useful internal representations of data. The approach has been widely used for basic cognition researches (e.g., memory, visual recognition) and practical applications. This paper, however, does not cite earlier work of the backpropagation method, such as the 1974 dissertation of Paul Werbos. Also in 1986, a book (Rumelhart & McClelland, 1986) described the creation of computer simulations of perception, giving to computer scientists their first testable models of neural processing, and which is now regarded as a central text in the field of cognitive science. Rumelhart's models of semantic cognition and specific knowledge in a diversity of learned domains using initially non-hierarchical neuron-like processing units continue to interest scientists in the fields of artificial intelligence, anthropology, information science, and decision science [16].

F. *Paul Werbos (1947-)*

Paul Werbos is a social scientist and machine learning pioneer best known for his 1974 dissertation that first described the process of training artificial neural networks through the backpropagation of errors. He also was a pioneer of recurrent neural networks and a two-year Presidents of the International Neural Network Society. In 1995, he was awarded the IEEE Neural Network Pioneer Award for the discovery of backpropagation and other basic neural network learning frameworks such as Adaptive Dynamic Programming [20].

G. *Geoff Hinton (1947-)*

Hinton is a cognitive psychologist and computer scientist, most noted for his work on neural networks, and especially for restricted Boltzmann machines stacked as deep-belief networks. Hinton was co-author of a highly-cited paper published in 1986 that popularized the backpropagation algorithm for training multi-layer neural networks, although they were not the first to discover the algorithm. Hinton is viewed by some as a leading figure in the deep learning community and is referred to by some as the "Godfather of Deep Learning." The dramatic image-recognition milestone of the AlexNet designed by his student Alex Krizhevsky for the Imagenet challenge 2012 helped to revolutionize the field of computer vision. Hinton was awarded the 2018 Turing Prize alongside Yoshua Bengio and Yann LeCun for their work on deep learning [17].

H. *Yann LeCun (1960-)*

LeCun is a computer scientist working in the fields of machine learning, computer vision, mobile robotics, and computational neuroscience. He is a professor at the Courant Institute of Mathematical Sciences at New York University, and Vice President, Chief AI Scientist at Facebook. He is well known for his work on optical character recognition and computer vision using convolutional neural networks (CNN), and is considered a founding father of convolutional nets. He is co-recipient of the 2018 ACM A.M. Turing Award for his work in deep learning [17].

I. *Kai-Fu Lee (1961-)*

Lee is a venture capitalist, technology executive, writer, and an artificial intelligence expert, and is currently based in Beijing, China. Lee developed the world's first speaker-independent, continuous speech recognition system as his Ph.D. thesis at Carnegie Mellon. He later worked as an executive, first at Apple, then SGI, Microsoft, and Google. He became the focus of a 2005 legal dispute between Google and Microsoft, his former employer, due to a one-year non-compete agreement that he signed with Microsoft in 2000 when he became its corporate vice president of interactive services. As a prominent figure in the Chinese internet sector, he was the founding director of Microsoft Research Asia, serving from 1998 to 2000, and president of Google China, serving from 2005 to 2009 [8].

J. *Andrew Ng (1963-)*

Ng is a computer scientist and one of the most prolific researchers in machine learning and AI. His work helped incite the recent revolution in deep learning. As a business executive and investor, Ng co-founded and led Google Brain and was a former Vice President and Chief Scientist at Baidu, building the company's Artificial Intelligence Group into a team of several thousand people. Ng teaches at Stanford University and was formerly Director of its AI Lab. Also a pioneer in online education, Ng co-founded Coursera and deeplearning.ai [18].

K. *Yashua Bengio (1964-)*

Bengio is a computer scientist, most noted for his work on neural networks and deep learning. He was a co-recipient of the 2018 ACM A.M. Turing Award for his work in deep learning. He is a professor at the Department of Computer Science and Operations Research at the Université de Montréal and scientific director of the Montreal Institute for Learning Algorithms. His research team is behind Theano [17].

L. *Jürgen Schmidhuber (1976-)*

Schmidhuber is a computer scientist noted for his work in the field of artificial intelligence, deep learning and neural networks. He is a co-director of the Dalle Molle Institute for Artificial Intelligence Research in Manno, Switzerland. He is sometimes called the "father of modern AI" or "father of deep learning." Together with his students he has published increasingly sophisticated versions of a type of recurrent neural network called the long short-term memory [21].

IV. Fathers of Deep Learning

So who are the fathers of deep learning? First consider that Hinton, LeCun and Bengio have just received the Turing Award as the fathers of the deep learning *revolution* "for conceptual and engineering breakthroughs that have made deep neural networks a critical component of computing" [17]. "Working independently and together, Hinton, LeCun and Bengio developed conceptual foundations for the field, identified surprising phenomena through experiments, and contributed engineering advances that demonstrated the practical advantages of deep neural networks. In recent years, deep learning methods have been responsible for astonishing breakthroughs in computer vision, speech recognition, natural language processing, and robotics—among other applications." [17].

The Turing Award recognized Hinton, LeCun and Bengio for the *revolution* caused by the revolutionary applications of deep learning networks spurred by their contributions primarily over the last ten or so years. We have evaluated several other candidates born after 1940 but Hinton, LeCun and Bengio are the outstanding candidates. However, is there a candidate that created and researched earlier neural networks having characteristics similar to today's networks? After careful review of the older candidates, we feel that Rosenblatt should be considered the elder father of deep learning, and the remainder of this section reviews his contributions.

Rosenblatt developed neural networks called perceptrons, probabilistic models for information, storage, and organization in the brain with the key properties of association or learning, generalization to new patterns, distributed memory, and biologically plausible brain models [13]. Most of the following figures were taken directly from his book. As was his preference, we will not capitalize the word "perceptron." His basic topological structure of the nervous system and its sources of information is shown in Fig. 4.

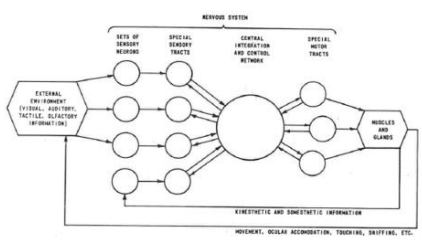

Fig. 4. Basic topological structure of the nervous system and its sources of information [13].

His definitions for the various types of perceptrons are:

- A perceptron is a network of sensory (S), association (A), and response (R) units with an interaction matrix of connection coefficients for all pairs of units
- A series-coupled perceptron is feed-forward S→A→R neural network
- A cross-coupled perceptron is a system in which some connections join units in the same layer
- A back-coupled perceptron is a system in which some connections flows back to an earlier layer

A simple perceptron is series-coupled with one R-unit and fixed S→A connections (Fig. 5). Note that this simple perceptron is a three-layer neural network with one hidden (A-unit) layer.

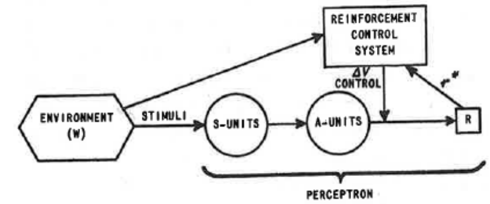

Fig. 5. Simple perceptron experimental system [13].

His general perceptron is a multi-layer perceptron (MLP) with a network of hidden (A-unit) layers (Fig. 6).

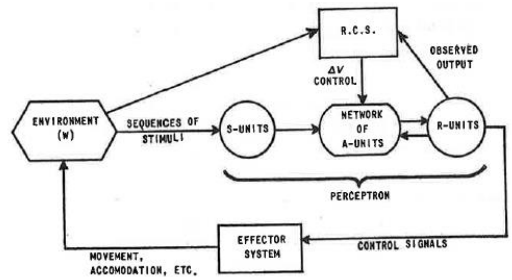

Fig. 6. General perceptron experimental system [13].

The first electronic perceptron, the Mark-I Perceptron (Fig. 7), is now located at the Smithsonian Institution.

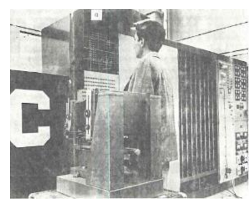

Fig. 7. Mark-I Perceptron [11].

The Mark-I Perceptron is a visual system model and classifier, a three-layer perceptron with fixed S→A and variable A→R connections (Fig. 8). The sensory (input) S-layer consists of 400 photosensitive units in a 20x20 grid, modeling a small retina. Connections from the input to the association layer were altered through plug-board wiring, but once wired they were fixed for the duration of an experiment. The association (hidden) A-layer consisted of 512 units (stepping motors), each of which could take several excitatory and inhibitory inputs. Connections from the association to the output layer were variable weights (motor-driven potentiometers) adjusted through the error-propagating training process. The response (output) R-layer consisted of 8 units.

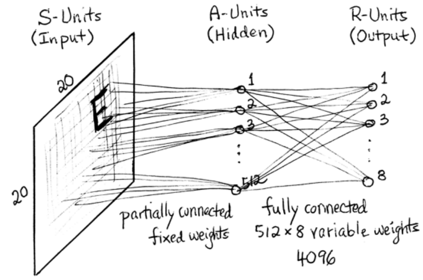

Fig. 8. Mark-I Perceptron (drawn by author).

The second electronic perceptron was the Tobermory Perceptron, an auditory system model and pattern classifier named after the talking cat, Tobermory, from a story by H.H. Munro (aka Saki). It was a large machine at that time (Fig. 9).

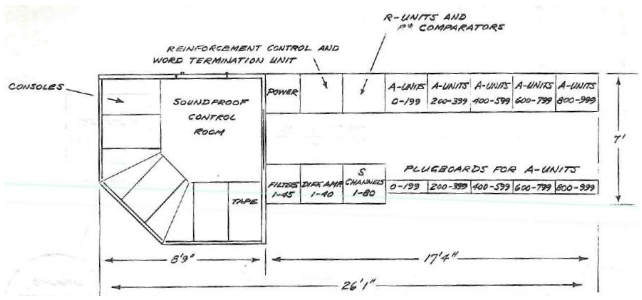

Fig. 9. Tobermory Perceptron floor plan [11].

The Tobermory machine consisted of:

- S-units: 45 band-pass filters and 80 difference detectors
- A1-units: 1600 (20 time samples from the 80 detectors)
- A2-units: 1000
- R-units: 12, with 12,000 adaptive weights A2→R-units.

Hardware implementations made good demonstrations but software simulations were far more flexible. In the 1960s Rosenblatt conducted many perceptron computer simulations. These computer simulations required machine language coding for speed and memory usage. A simulation software package was developed so a user could specify the number of layers, the number of units per layer, type of connections between layers, etc. Computer time was used at Cornell and at NYU.

Rosenblatt's *Principles of Neurodynamics* [13] has four parts. Part I was an historical review of brain modeling approaches, physiological and psychological considerations, and basic definitions and concepts of the perceptron approach. Part II discussed the three-layer, series-coupled perceptrons – the mathematical underpinnings and the experimental results. The mathematical underpinnings included several theorems:

- Convergence Theorem: Given a simple perceptron, a stimulus world W, and any classification C(W) for which a solution exists, then if all stimuli in W re-occur in finite time, the error correction procedure will always find a solution. Perceptrons were the first neural networks that could learn the weights!
- Solution Existence Theorem: The class of simple perceptrons for which a solution exists to every classification C(W) of possible environments W is non-empty. That is, there exists an S→A→R (minimum of three layers) feedforward perceptron that can solve any classification problem.

Part III covered multi-layer (more than three layers) and cross-coupled perceptrons, and Part VI back-coupled perceptrons. Rosenblatt used the book to teach an interdisciplinary course "Theory of Brain Mechanisms" that drew students from Cornell's Engineering and Liberal Arts colleges.

In Rosenblatt's MLPs the first hidden (A1) layer usually consisted of biologically-plausible detector units like edge and line detectors for visual image classification [7] (Fig. 10) and the second (A2) hidden layer consisted of various combinations of A1 units to detect special features or motion (Fig. 11).

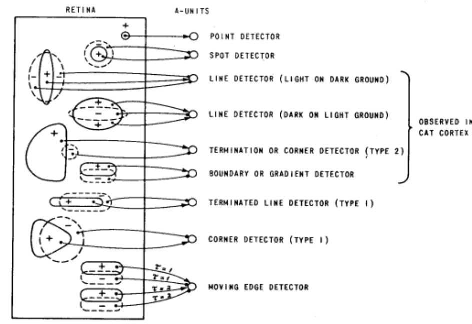

Fig. 10. Biologically plausible A1-units: broken lines indicate inhibitory fields, solid lines excitatory fields [13].

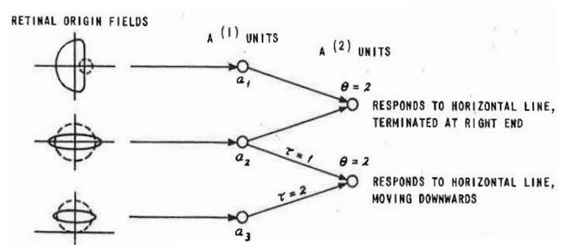

Fig. 11. Biologically plausible detector A2-units [13].

Rosenblatt studied three and four-layer series-coupled perceptrons with two sets of variable weights but was unable to find a suitable training procedure like back-propagation to train the whole system (Fig. 12). Rosenblatt also studied cross-coupled perceptrons in which some connections join units of the same type (S, A, and/or R), see Fig. 13.

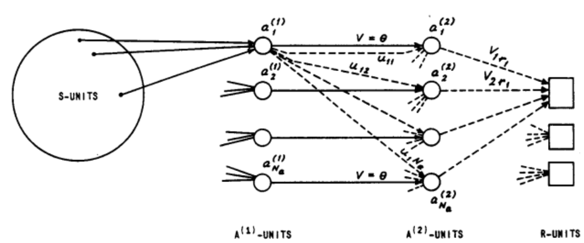

Fig. 12. Perceptron with two variable weight sets (dotted lines) [13].

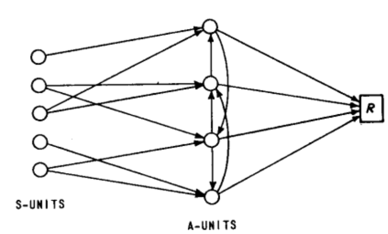

Fig. 13. Cross-coupled perceptron [13].

And he studied back-coupled perceptrons with feedback paths from units located near the output end of the system to units closer to the input end (Fig. 14).

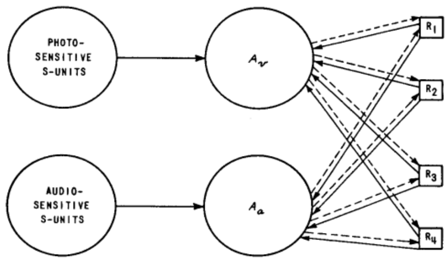

Fig. 14. A back-coupled perceptron [13].

Frank Rosenblatt and Marvin Minsky debated at conferences the value of biologically inspired computation, Rosenblatt arguing that his neural networks could do almost anything and Minsky countering that they could do little. Minsky, wanting to direct government funding away from neural networks and towards his own areas of interest, collaborated with Seymour Papert to publish *Perceptrons* [10], where they asserted about perceptrons (page 4), "Most of this writing ... is without scientific value..." Minsky, although well aware that powerful perceptrons have multiple layers and even Rosenblatt's basic feed-forward perceptrons have three layers, defined a perceptron as a two-layer machine that can handle only linearly separable problems and, for example, cannot solve the exclusive-OR problem. The book, unfortunately, stopped government funding in neural networks and precipitated an "AI Winter" that lasted about 15 years. This lack of funding also ended Rosenblatt's research in neural networks.

It is clear that Rosenblatt had all the basically ingredients of the deep learning systems of today, except most notably the backpropagation training algorithm, although he was the first to develop a training algorithm for one set of weights that was guaranteed to find a solution if one existed. The award-winning, deep learning systems are multilayer feed-forward, Convolutional Neural Networks (CNNs). These are just multilayer perceptrons with fixed, biologically-inspired connections in the early layers, and it is interesting that they are generally called multilayer perceptrons (MLP). The Recurrent Neural Networks (RNNs) have connections within the same layer and/or connections back to earlier layers which are just Rosenblatt's cross-coupled and back-coupled perceptrons.

Although conceptually the same as perceptrons, the deep learning networks have the advantages of:

- Increased computing power + GPUs (50 years from 1960s)
- Massive training data (big data) not available in 1960s
- Backpropagation algorithm not available in 1960s that allows training of the complete system, pulling everything together

V. Conclusions

This paper evaluated a number of candidates as potential fathers of the deep learning technology. We conclude that Rosenblatt developed and explored all of the basic ingredients of the deep learning systems of today. There are two additional factors that perhaps should be taken into account. The first is the occurrence of the "AI Winter" triggered by Minsky's book in 1969 that stopped government funding in neural networks and ended Rosenblatt's research on perceptrons when he was only 41 years old. The second factor is that Rosenblatt was not alive to take up his research on perceptrons when work resumed on neural networks in the 1980s because he died in a tragic boating accident in 1971. Had he been able to continue his work on perceptrons there is no telling how far he could have pushed the technology in those early years.

Because Rosenblatt developed and explored all of the basic ingredients of the deep learning systems of today, we believe he should be recognized as the Father of Deep Learning, or at least be considered one of the fathers together with Hinton, LeCun and Bengio.

References

1. Abraham, T.H. (2016). *Rebel genius: Warren S. McCulloch's transdisciplinary life in science*, M.I.T. Press.
2. AI Venn Diagram (2016). https://whatsthebigdata.com/2016/10/17/visually-linking-ai-machine-learning-deep-learning-big-data-and-data-science/
3. Asaro, P. M. (2008). "From Mechanisms of Adaptation to Intelligence Amplifiers: The Philosophy of W. Ross Ashby," in Michael Wheeler, Philip Husbands and Owen Holland (eds.) The Mechanical Mind in History, Cambridge, Massachusetts: MIT Press, pp. 149–184.
4. Brown, R.M.; Milner, P.M. (2003). "The Legacy of Donald O. Hebb: More than the Hebb Synapse". Nature Reviews Neuroscience. **4** (12): 1013–1019.
5. Goodfellow, I., Bengio, Y. & Courville, A. (2016). Deep Learning, M.I.T. Press, 2016.
6. Grigsby, S. (2018). Artificial Intelligence for Advanced Human-Machine Symbiosis. 10.1007/978-3-319-91470-1_22.
7. Hubel, D.H. & Wiesel, T.N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology. **124** (3): 574–591.*
8. Kai-Fu Lee named Asia House Asian Business Leader (2018). Asia House. Retrieved 2019.
9. Krizhevsky, A., Sutskever, I. & Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems 25 (NIPS 2012).
10. Minsky, M. & Papert, S. (1969). *Perceptrons: An Introduction to Computational Geometry*, MIT Press, 1969
11. Nagy, G. (2011). Frank Rosenblatt, my distinguished advisor. Keynote address, Research Day Conference, Pace University, 2011.
12. Parloff, R. (2016). Why Deep Learning is Suddenly Changing Your Life, *Fortune*, Sept, 2016.
13. Rosenblatt, R. (1962). *Principles of Neurodynamics*, Spartan, 1962.
14. Rumelhart, D.E., Hinton, G.E. & Williams, R.J. (1986). Learning representations by back-propagating errors. *Nature* **323** (6088): 533–536.
15. Rumelhart, D.E. & McClelland, J.L. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition,* MIT Press.
16. The Rumelhart Prize (2013). David E. Rumelhart: A Scientific Biography. http://rumelhartprize.org/?page_id=10
17. Turing Award (2019). Fathers of the Deep Learning Revolution Receive ACM A. M. Turing Award. https://www.acm.org/media-center/2019/march/turing-award-2018.
18. Weinberger, M. (2018). One of the world's most famous computer scientists reveals his 'playbook' for bringing AI to every business. *Business Insider,* 2018.
19. Wikipedia-List (2019), https://en.wikipedia.org/wiki/List_of_people_considered_father_or_mother_of_a_scientific_field, Accesed 2019.
20. Wikipedia-Werbos (2019), https://en.wikipedia.org/wiki/Paul_Werbos. Accessed 2019.
21. Wong, Andrew (2018). The 'father of A.I' urges humans not to fear the technology. CNBC. Accessed 2019.