State of the Art Artificial Neural Network, Deep Learning, and the Future Generation

** Munef Abdullah Ahmed*
*** Stefan Trausan-Matu*

Abstract

The use of neural networks or artificial intelligence or deep learning in its broadest and most controversial sense has been a tumultuous journey involving three distinct hype cycles and a history dating back to the 1960s. Resurgent and enthusiastic interest in machine learning and its applications bolster the case for machine learning as a fundamental computational kernel. Furthermore, researchers have demonstrated that machine learning can be utilized as an auxiliary component of applications to enhance or enable new types of computation such as approximate computing or automatic parallelization. In our view, machine learning becomes not the underlying application, but a ubiquitous component of applications. In recent years, deep learning in artificial neural networks (ANN) has won numerous contests in pattern recognition and machine learning; this view necessitates a different approach towards the deployment of ANN and deep learning.

Keywords—Artificial neural network, deep learning, machine learning, neural network.

I. INTRODUCTION

Most times neural network architectures mimic the human mind and attempt to benefit from the features of nerve cells as much as possible. Deep neural networks (DNNs), like other neural networks, mostly use identical features rather than mimicking the human brain in making precise measurements. Deep learning (DL) is also known as deep machine learning, hierarchical learning or deep structured learning. It is used for the processing of nonlinear information using unsupervised or supervised features for classification and pattern analysis [1]. DLNNs with several layers are considered beneficial since the neural framework can be trained in the first layer to learn parts of the problem, while the neurons are trained in the second layer to learn a different part of the problem; as well as the third hidden layer and so on. So, DLNNs are considered more useful than the shallow networks. A DNN is an ANN which has input layer units and several hidden input and output layers and can be used to model and solve several complex nonlinear relationships. The feature of reducing layers can be configured by adding extra layers; and with a fewer number of units, can model difficult data with the same efficiency as the shallow networks [2]. Deep learning is a more effective approach than the traditional neural network in solving tasks that involve the availability of data [3]. It is used for solving problems on the MIC architecture without changing the baseline algorithm [4]. It is effective in facial expression recognition [5] and in solving many data stream mining tasks [6].

To fully understand the motivation behind this review, this paper is organized into sections as follows: the introduction has been made in the first section; section II presents the motivation for the study, while section III presents the recurring themes of deep learning (DL).In section IV, the supervised neural networks (NNs) is presented, while section V presents the unsupervised learning (UL).Section VI reviews the paradigms of learning; section VII presents the recurrent NNs; section VII presents backpropagation, and section IX discusses self-organizing networks. A conclusion from the review is presented in section X, which is followed by references.

II. THE MOTIVATION

Recently, DNNs have won several awards in the field of pattern recognition due to its flexibility and ability to

Manuscript received September 29, 2017; revised October 14, 2017; accepted October 15, 2017. Date of publication November 6, 2017.
M. A. Ahmed is with Splaiul Independenței 313, București, Romania - 060042. (email : yabrahmun@yahoo.com)
Ș. Trâușan-Matu is Professor , Faculty of Automatic Control and Computers, University Polytechnic of Bucharest
Splaiul Independenței 313, București, Romania-060042. (email: stefan.trausan@cs.pub.ro)
DOI: 10.17010/ijcs/2017/v2/i6/120440

solve many types of problems that cannot be solved using normal computation methods. It is widely used in the field of image and video processing and has shown amazing results in natural language processing. recurring themes of deep learning (DL) (Dynamic Programming (DP)): This is important in the presence of deterministic assumptions because it simplifies credit tasks [7]. Backpropagation is an example DP used in supervised learning. Hidden Markov Models (HMMs) are used in DP algorithms [8], and they are important when the neural network system contains additional ideas or concepts such as graphical models [9], [10].

1) Minimum Description Length (MDL): This means that the complexity of finding a solution by depending on the length of a program. The shortest programs are used in computing [11], [12]; some consider the program runtime [13]- [15] while others consider only constant runtime of the program [16]. MDL has been proposed to regularize and reduce the weight complexity of NNs [17].

2) Graphics Processing Units (GPU): There have been several previous attempts to fasten the GPUs used in NNs [18], [19], though a standard hardware has been introduced [20]. In recent years, GPUs have been widely used in video pl)ys and self-drive cars, leading to and reduction in their prices and improvement of their features in the aspects of training and speed [21], [22].

III. SUPERVISED NNS

1) Group Method of Data Handling (GMDH): This is a method used for training several NNs; some sources consider this method as the first that used multi-layer perception in deep learning system [23],[24],[25]. The polynomial activation functions may be used in the units of GMDH because this form of network has many applications [26],[27].

2) Convolutional Neural Networks (CNN): The CNN was first introduced by Fukushima [28] [29], [30], [31] who set the weights using spatial averaging (an unsupervised learning rule) [30], [31], instead of max-pooling.

IV. UNSUPERVISED LEARNING (UL)

Previously, several researchers have focused on learning without a teacher; among these researchers [32] [33], many have used several UL methods in the field of information theory [34]. Furthermore, UL has been proposed in automatic distribution [35], autoencoder, and linguistic structure encoding [36].

V. PARADIGMS OF LEARNING

Neural networks can be trained by providing it with a range of teaching patterns. These patterns should be carefully selected because they contribute to the speed of the network training, and by using some learning rules, the neural network can either change its training weight and be trained by the set weights.

1) Supervised learning: In supervised learning (SL), there are matching output patterns for each example offered to the neural network. Thus, the neural network can adjust its weight by knowing the difference between the current director and its goal [37], [38]. The most famous learning algorithms use supervised learning. Backpropagation [39], [40], [41], Delta rule, and supervised learning have many possible applications, as in speech recognition, automatic translation, hand-writing recognition, and stock market prediction [42], [43].

2) Unsupervised Learning: The unsupervised learning (UL) is the second type of learning paradigm used in classification networks.

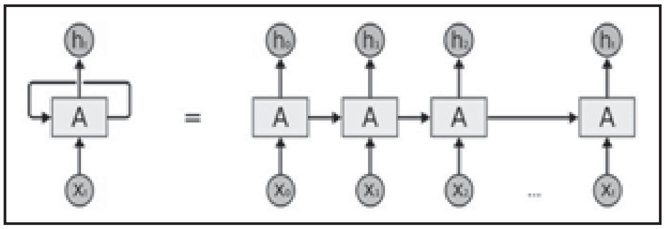
VI. RECURRENT NEURAL NETWORKS

Recurrent Neural Networks (RNNs) are considered as one of the fields of artificial neural networks similar to feedforward neural networks. The difference between them is that the forward data flow in the feedforward neural networks has no feedback nor cycle, but in the RNNs, there is at least one feedback. We can connect a hidden unit to the input units, itself orconnect all units with each other; thus, the internal state of the network enables it to learn sequences.

RNNs are powerful because of the distributed hidden state that makes it possible for them to efficiently retail their prior knowledge and the nonlinear dynamics that enables them to update their hidden state in intricate ways. There are several types of recurrent neural networks architectures. Fig. 1. illustrates an unrolled recurrent network, where A is a chunk of NN, Xj is the jth input of the NN, and h is the output at step j. The RNN is powerful in the fieldof speech [52] and handwriting recognition [53].

The Long Short-Term Memory (LSTM) is one of the important types of recurrent neural network architectures proposed by Jurgen Schmidhuber and Sepp Hoch Reiter in 1997 [54], [55].There are several types of LSTM that offer good solutions to problem-specific topologies [56] and in modifying self-connection [57]. Several tasks are solved by learning the rule of languages explained by

Fig. 1. An unrolled Recurrent Neural Network



deterministic finite state automata [58]-[60]. Free languages which cannot be represented by any other method [61], [62] can be handled with RNN [63], [64]. The result of HMM-based systems is comparable to that of LSTM [65]-[66]. LSTM is used in speech recognition [67], robot localization [68], online drive distraction detection [69], and robot control [70]. Other applications include voice detection [30], handwriting recognition [71], language identification [72], audio onset detection [64], machine translation [73], and social signal classification [74].

The depth of a feed-forward RNN does not apply to a recurrent RNN because it does not depend only on the number of layers in the network which must be multiple nonlinear layers but also because of its temporal structure. There are several types of depth of a recurrent neural network- from the input-to-hidden function and deep hidden-to-hidden function to the hidden-to-output function and stack of hidden states [75].

VII. BACKPROPAGATION

Backpropagation (BP) was first mentioned in a thesis submitted in 1970 by Linnainmaa, and maybe the first touch on the effectiveness of backpropagation [76],[77]. Control parameters were adapted by backpropagation, which is a way for multilayer thresholds using backpropagation [78]. A particular application for backpropagation in neural networks [79] also observed an increasing use of backpropagation for research [80], [81]. In convolutional NNs, [82], [83], backpropagation helped to improve the use of NNs to distinguish fingerprints [84]. Several types of researchers have used backpropagation in neural networks in the field of engineering [85], marketing [86], and medicine [87].

Back-propagation (BP) is an important area of NNs in which BP can be defined as a learning algorithm to reduce the error function. In BP, the errors in the neurons cannot be seen but can be known in the output layers. These errors can be calculated by finding the difference between the output and the desired output. The error in the hidden layers is also updated with the other parameters by depending on weight changes in the layer ahead. BP is a good method for training a multi-layered NN, and this is the motivation for developing BP [88].The NN is used to reduce the error in the gradient descent [89] where there are nonlinear, complex, differentiable and multi-layer parameters [90]-[92]; efficient error back propagation was first reported in a master thesis [76] by using adapting control parameters to minimize the cost functions [93]. A multi-layer threshold NNs [78] was first used with BP in computer programs for finding automatic implementation [94]. The application of BP in RNNs was reported by [95]-[100], and in convolutional Nns LeCun, Boser, Denker, Henderson, Howard, Hubbard, and Jackel [82] and Mohammed, Salih, Ţăpuş, and Hasan[101].

The advantages of BP are in its versatility and accuracy; however, it suffers from problems such as the length of time to resolve, not without complexity. There are many BP applications in medicine, bioinformatics, marketing, and others. BP has a high ability to solve problems that cannot be solved using traditional machine learning [39].

VIII. SELF-ORGANIZING NETWORKS

When we consider biological neural networks, several thoughts may come to mind; one is that our brain receives a lot of information ona daily basis and can recall impressions. The brain does not have any guidance, desired output or preparatory tests, but our cerebrum reacts to the external inputs by changing the state. Based on this basic thought of the biological neural networks and how they connect to the external world and organize themselves, Professor Teuvo Kohonen in the 80s, developed the self-organizing maps (SOM) which learn without a teacher, completely unsupervised [51], and hence, SOM is a type of artificial neural network (ANN).There are several types of self-organizing networks that have been used in a wide range of applications; Rumelhart and Zipser proposed the most basic schemes of SOM, which is a competitive learning algorithm [80].The earliest applications of SOM were in medicine, speech recognition, video and image processing, text or document mining, robotics, hardware implementation, and in computer animation.

IX. CONCLUSION

Neural Networks and Deep Neural Networks can be used for solving many problems that cannot be solved by any other method. Supervised and unsupervised learning have a strong relevance with deep learning in neural networks. GPU and convolutional NNs won competitions not only in pattern recognition but also in other fields like object detection and image segmentation. Backpropagation is used in training Neural Networks that have multilayers; hence, there is an increase in the research interest in this area of study. The Long Short-Term Memory is one of the important types of recurrent neural network architecture. As we know that the main task for recurrent neural networks is to use sequential information, with the output being based on prior computations. Another strategy to be considered about recurrent neural networks is that they have a memory that can capture previous calculations.

On the basis of the basic thought of the biological neural networks and how they are connected to the external world and organized, it becomes certain that self-organizing maps which are completely unsupervised can open a wide space for the development and use of this line of thought in several applications.

REFERENCES

[1] L. Deng, and D. Yu, "Deep learning: Methods and applications," *Foundations and Trends® in Signal Processing*, vol. 7, no. 3-4, pp. 197-387, 2014. doi: http://dx.doi.org/10.1561/2000000039

[2] Y. Bengio, "Learning deep architectures for AI," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1-127, 2009.

[3] Y. Wang, M. Liu, and Z. Bao, "Deep learning neural network for power system fault diagnosis," in the *Control Conference (CCC), 2016 35th Chin., Chengdu,* 2016, pp. 6678-6683. doi: 10.1109/ChiCC.2016.7554408

[4] E. Milova, S. Sveshnikova, and I. Gankevich, "Speedup of deep neural network learning on the MIC-architecture," in *High Performance Computing & Simulation (HPCS), 2016 International Conference*, pp. 989-992. doi: 10.1109/HPCSim.2016.7568443

[5] Y. Guo, D. Tao, J. Yu, H. Xiong, Y. Li, and D. Tao, "Deep Neural Networks with Relativity Learning for facial expression recognition," in *Multimedia & Expo Workshops (ICMEW), 2016 IEEE Int. Conf.*, pp. 1-6. doi: 10.1109/ICMEW.2016.7574736

[6] Y. Bodyanskiy, O. Vynokurova, I. Pliss, G. Setlak, and P. Mulesa, "Fast learning algorithm for deep evolving GMDH-SVM neural network in data stream mining tasks," in *Data Stream Mining & Processing (DSMP), IEEE First Int. Conf.*, pp. 257-262. doi: 10.1109/DSMP.2016.7583555

[7] R. Bellman, *Dynamic Programming*. Princeton University Press, Princeton, 1957.

[8] R. L. Stratonovich, "Conditional markov processes," *Theory of Probability & its Appl.*, vol. 5, no. 2, pp. 156-178, 1960. doi: https://doi.org/10.1137/1105015

[9] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. of the Roy. Statistical Soc*. Series B (methodological), vol. 39, no. 1, pp. 1-38, 1977.

[10] P. Baldi, and Y. Chauvin, "Hybrid modeling, HMM/NN architectures, and protein applications," Neural Computation, vol. 8, no. 7, pp. 1541-1565, 1996. doi: 10.1162/neco.1996.8.7.1541

[11] A. N. Kolmogorov, "Three approaches to the quantitative definition of information," *Problems of Inform. Transmission*, vol. 1, no. 1, pp. 1-7, 1965. doi: http://dx.doi.org/10.1080/00207166808803030

[12] P. D. Grünwald, I. J. Myung, and M. A. Pitt, Advances in Minimum Description Length: Theory and Appl. MIT Press, 2005.

[13] E. Allender, "Applications of time-bounded Kolmogorov complexity in complexity theory," Kolmogorov complexity and computational complexity. Springer, 1992, pp. 4-22.

[14] J. Schmidhuber, "The speed prior: a new simplicity measure yielding near-optimal computable predictions," Proceedings of the 15th Annu. Conf. on Computational Learning Theory (COLT 2002), Sydney, Australia, Lecture Notes in Artificial Intelligence. Springer, pp. 216-228.

[15] J. Schmidhuber, "Draft: Deep Learning in Neural Networks: An Overview," *Neural Networks*, vol. 61, pp. 85-117, 2014.

[16] J. Rissanen, "Stochastic complexity and modeling," *The Ann. of Statist.*, vol. 14, no. 3, pp. 1080-1100, 1986.

[17] M. C. P. de Souto, C. Marc'ilio, and W. R. De Oliveira, "The loading problem for pyramidal neural networks".

[18] L. D. Jackel, B. Boser, H. Graf, J. S. Denker, Y. Le Cun, D. Henderson, O. Matan, R. Howard, and K. Baird, "VLSI implementations of electron. neural networks: An example in character recognition," in *Syst., Man and Cybernetics 1990. Conf. Proc.*, pp. 320-322. doi: 10.1109/ICSMC.1990.142119

[19] B. Widrow, D. E. Rumelhart, and M. A. Lehr, "Neural networks: Appl. in industry, business, and sci.," *Commun. of the ACM*, vol. 37, no. 3, pp. 93-106, 1994.

[20] D. Anguita, and B. A. Gomes, "Mixing floating-and fixed-point formats for neural network learning on neuroprocessors," *Microprocessing and*

Microprogramming, vol. 41, no. 10, pp. 757-769, 1996. doi: https://doi.org/10.1016/0165-6074(96)00012-9
[21] D. Cire?an, U. Meier, J. Masci, and J. Schmidhuber, "A committee of neural networks for traffic sign classification," in Neural Networks (IJCNN), The 2011 *Int. Joint Conf.*, pp. 1918-1921. doi: 10.1109/IJCNN.2011.6033458.
[22] E. D. Dickmanns, R. Behringer, D. Dickmanns, T. Hildebrandt, M. Maurer, F. Thomanek, and J. Schiehlen," The seeing passenger car' VaMoRs-P'," in *Proc. of the Intelligent Vehicles '94 Symp.,* pp. 68-73. doi: 10.1109/IVS.1994.639472
[23] A. G. Ivakhnenko, and V. G. Lapa, "Cybernetic predicting devices," DTIC Document, 1966.
[24] A. G. Ivakhnenko, "The group method of data handling-a rival of the method of stochastic approximation," *Soviet Automat. Control*, vol. 1, no. 3, pp. 43-55, 1968.
[25] A. G. Ivakhnenko, "Polynomial theory of complex syst.," *IEEE transactions on Systems, Man, and Cybernetics,* vol. SMC - 1, no. 4, pp. 364-378, 1971. doi: 10.1109/TSMC.1971.4308320
[26] A. Ivakhnenko, and G. Ivakhnenko, "The rev. of problems solvable by algorithms of the group method of data handling (GMDH)," *Pattern Recognition and Image Anal. C/C Of Raspoznavaniye Obrazov I Analiz Izobrazhenii*, vol. 5, no. 4, pp. 527-535, 1995.
[27] T. Kondo, J. Ueno, and S. Takao, "Multi-layered GMDH-type neural network and self-selecting optimum neural network architecture and its application to 3-dimensional medical image recognition of blood vessels," *Int. J. of Innovative Computing Inform., and Control,* vol. 4, no. 1, pp. 175-187, 2008.
[28] K. Fukushima, "Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position-Neocognitron," *Electron. & Commun. Japan*, vol. 62, no. 10, pp. 11-18, 1979.
[29] K. Fukushima, "Neocognitron--A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position," *NHK*, no. 15, pp. 106-115, 1981.
[30] K. Fukushima, "Artificial vision by multi-layered neural networks: Neocognitron and its advances," *Neural Networks*, vol. 37, pp. 103-119, 2013. doi: https://doi.org/10.1016/j.neunet.2012.09.016
[31] K. Fukushima, "Increasing robustness against background noise: Visual pattern recognition by a neocognitron," *Neural Networks*, vol. 24, no. 7, pp. 767-778, 2011. doi: https://doi.org/10.1016/j.neunet.2011.03.017
[32] D. O. Hebb, "The Org. of behavior," New York: Wiley, 1949.
[33] M. A. Carreira-Perpinan, "Continuous latent variable models for dimensionality reduction and sequential data reconstruction," Ph. D. dissertation, *Dept. of Comput. Sci.*, University of Sheffield, United Kingdom, 2001.
[34] G. E. Hinton, P. Dayan, B. J. Frey, and R. M. Neal, "The wake-sleep algorithm for unsupervised neural networks," *Sci.*, vol. 268, no. 5214, pp. 1158-1161, 1995.
[35] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statist.," *IEEE Trans. on Signal Process*., vol. 45, no. 2, pp. 434-444, 1997. doi: 10.1109/78.554307
[36] O. Melnik, S. Levy, and J. Pollack, "RAAM for infinite context-free languages," in Neural Networks, 2000. IJCNN 2000, *Proc.* of the IEEE-INNS-ENNS Int. Joint Conf., 2000, pp. 585-590.
[37] J. E. Fieldsend, and R. M. Everson, "Multiobjective supervised learning," in Multiobjective Problem Solving from *Nature*, Springer, 2008, pp. 155-176.
[38] B. Schrauwen, D. Verstraeten, and J. Van Campenhout, "An overview of reservoir computing: Theory, Appl., and Implementations," in Proc. of the 15th Eur. Symp. on Artificial Neural Networks, pp/ 471-482.
[39] A. A. Hameed, B. Karlik, and M. S. Salman, "Back-propagation algorithm with variable adaptive momentum," Knowledge-Based Syst., vol. 114, pp. 79-87, 2016. doi: https://doi.org/10.1016/j.knosys.2016.10.001
[40] R. Karimi, F. Yousefi, M. Ghaedi, and K. Dashtian, "Back propagation artificial neural network and central composite design modeling of oper. parameter impact for sunset yellow and azur (II) adsorption onto MWCNT and MWCNT-Pd-NPs: Isotherm and kinetic study," *Chemometrics and Intelligent Laboratory Syst.*, vol. 159, pp. 127-137, 2016. doi: https://doi.org/10.1016/j.chemolab.2016.10.012
[41] L. Wang, Y. Zeng, and T. Chen, "Back propagation neural network with adaptive differential evolution algorithm for time series forecasting," *Expert Syst. with Appl.*, vol. 42, no. 2, pp. 855-863, 2015. doi: https://doi.org/10.1016/j.eswa.2014.08.018
[42] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *J. of Artificial Intell. Res.*, vol. 4, pp. 237-285, 1996.
[43] J. Schmidhuber, "Completely self-referential optimal reinforcement learners," *Artificial Neural Networks: Formal Models and their Appl.-ICAN*N 2005,

pp. 223-233, 2005.
[44] D. E. Rumelhart, D. Zipser, "Feature discovery by competitive learning," *Parallel Distributed Process.*, vol. 9, no. 1, pp. 75-112, 1985.
[45] N. Kambhatla, and T. K. Leen, "Dimension reduction by local principal component anal.," *Neural Computation*, vol. 9, no. 7, pp. 1493-1516, 1997. doi: 10.1162/neco.1997.9.7.1493
[46] B. Krose and P. Van der Smagt, *An Introduction to Neural Networks* (8th Ed.). University of Amsterdam, 1996.
[47] P. Foldiak, "Sparse coding in the primate cortex," in *The Handbook of Brain Theory and Neural Networks*, 2nd Ed. MIT Press, 2002.
[48] M. Klapper-Rybicka, N. N. Schraudolph, and J. Schmidhuber, *Unsupervised Learning in LSTM Recurrent Neural Networks*, pp. 684-691.
[49] M. Franzius, H. Sprekeler, and L. Wiskott, "Slowness and sparseness lead to place, head-direction, and spatial-view cells," *PLoS Computational Biology*, vol. 3, no. 8, pp. e166, 2007. doi: 10.1371/journal.pcbi.0030166
[50] S. Waydo, and C. Koch, "Unsupervised learning of individuals and categories from images," *Neural Computation*, vol. 20, no. 5, pp. 1165-1178, 2008.
[51] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Anal. and Mach. Intell.*, vol. 35, no. 8, pp. 1798-1828, 2013. doi: 10.1109/TPAMI.2013.50
[52] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," pp. 338-342.
[53] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber, "A novel connectionist system for unconstrained handwriting recognition," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, vol. 31, no. 5, pp. 855-868, 2009. doi: 10.1109/TPAMI.2008.137
[54] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
[55] J. A. Pérez-Ortiz, F. A. Gers, D. Eck, and J. Schmidhuber, "Kalman filters improve LSTM network performance in problems unsolvable by traditional recurrent nets," *Neural Networks*, vol. 16, no. 2, pp. 241-250, 2003.
[56] J. Bayer, D. Wierstra, J. Togelius, and J. Schmidhuber, "Evolving memory cell structures for sequence learning," *Artificial Neural Networks-ICANN 2009*, pp. 755-764.
[57] F. A. Gers, and J. Schmidhuber, "LSTM recurrent networks learn simple context-free and context-sensitive languages," *IEEE Trans. on Neural Networks*, vol. 12, no. 6, pp. 1333-1340, 2001. doi: 10.1109/72.963769
[58] W. M. Kistler, W. Gerstner, and J. L. van Hemmen, "Reduction of the Hodgkin-Huxley equations to a single-variable threshold model," *Neural Computation*, vol. 9, no. 5, pp. 1015-1045, 1997. doi: 10.1162/neco.1997.9.5.1015
[59] A. Vahid, and C. W. Omlin, "A mach. learning method for extracting symbolic knowledge from recurrent neural networks," *Neural Computation*, vol. 16, no. 1, pp. 59-71, 2004.
[60] Z. Zeng, R. M. Goodman, and P. Smyth, "Discrete recurrent neural networks for grammatical inference," *IEEE Trans. on Neural Networks*, vol. 5, no. 2, pp. 320-330, 1994. doi: 10.1109/72.279194
[61] G. Sun, C. L. Giles, H. Chen, and Y. Lee, *The Neural Network Pushdown Automaton: Model, Stack, and Learning Simulations*, 1998.
[62] P. Rodriguez, J. Wiles, and J. L. Elman, "A recurrent neural network that learns to count," Connection Sci., vol. 11, no. 1, pp. 5-40, 1999.
[63] J. Martens, and I. Sutskever, "Learning recurrent neural networks with Hessian-free optimization," in *Proc. of the 28th Int. Conf. on Mach. Learning*, Bellevue, WA, USA, 2011, pp. 1033-1040.
[64] E. Marchi, G. Ferroni, F. Eyben, L. Gabrielli, S. Squartini, and B. Schuller, "Multi-resolution linear prediction based features for audio onset detection with bidirectional LSTM neural networks," in *IEEE Int. Conf. Acoustics.*, Speech and Signal Process. (ICASSP), 2014, pp. 2164-2168.
[65] G. Deco, and E. T. Rolls, "Neurodynamics of biased competition and cooperation for attention: A model with spiking neurons," *J. of Neurophysiology*, vol. 94, no. 1, pp. 295-313, 2005. doi: 10.1152/jn.01095.2004
[66] M. Woellmer, B. Schuller, and G. Rigoll, "Keyword spotting exploiting long short-term memory," *Speech Communication*, vol. 55, no. 2, pp. 252-265, 2013. doi: https://doi.org/10.1016/j.specom.2012.08.006
[67] J. Schmidhuber, "My first deep learning system of 1991+ deep learning timeline 1962-2013," arXiv preprint arXiv:1312.5548, 2013.
[68] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks," in *Proc. ICML '06 Proc. of the 23rd Int. Conf. on Mach. Learning*, pp. 369-376.
[69] M. Wollmer, C. Blaschke, T. Schindl, B. Schuller, B.

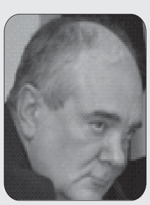
Farber, S. Mayer, and B. Trefflich, "Online driver distraction detection using long-short-term memory," *IEEE Trans. on Intelligent Transportation Syst*., vol. 12, no. 2, pp. 574-582, 2011. doi: 10.1109/TITS.2011.2119483
[70] H. Mayer, F. Gomez, D. Wierstra, I. Nagy, A. Knoll, and J. Schmidhuber, "A system for robotic heart surgery that learns to tie knots using recurrent neural networks," *Advanced Robotics*, vol. 22, no. 13-14, pp. 1521-1537, 2008.doi: https://doi.org/10.1163/156855308X360604
[71] A. Graves, M. Liwicki, H. Bunke, J. Schmidhuber, and S. Fernández, "Unconstrained on-line handwriting recognition with recurrent neural networks," in *Advances in Neural Inform. Process. Syst. 20 (NIPS 2007)*, pp. 577-584.
[72] I. J. Goodfellow, M. Mirza, D. Xiao, A. Courville, and Y. Bengio, "An empirical investigation of catastrophic forgetting in gradient-based neural networks," arXiv preprint arXiv:1312.6211, 2013.
[73] H. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *NIPS'14 Proc. of the 27th Int. Conf. on Neural Inform. Process. Syst.* - vol. 2, pp. 3104-3112.
[74] R. Brueckner and B. Schulter, "Social signal classification using deep BLSTM recurrent neural networks," 2014 *IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, pp. 4823-4827. doi: 10.1109/ICASSP.2014.6854518
[75] R. Pascanu, C. Gulcehre, K. Cho, and Y. Bengio, "How to construct deep recurrent neural networks," arXiv:1312.6026 [cs.NE], 2013.
[76] S. Linnainmaa, "The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors," Master's Thesis (in Finnish), Univ. Helsinki, pp. 6-7, 1970.
[77] S. Linnainmaa, "Taylor expansion of the accumulated rounding error," *BIT Numerical Mathematics*, vol. 16, no. 2, pp. 146-160, 1976.
[78] L. Bobrowski, "Learning processes in multilayer threshold nets," *Biological Cybern.*, vol. 31, no. 1, pp. 1-6, 1978.
[79] P. J. Werbos, "Appl. of advances in nonlinear sensitivity analy.," *System Modeling and Optimization*, pp. 762-770: Springer, berlin, 1982.
[80] D. E. Rumelhart, and D. Zipser, "Feature discovery by competitive learning," *Cognitive Sci.*, vol. 9, no. 1, pp. 75-112, 1985. doi: https://doi.org/10.1016/S0364-0213(85)80010-0
[81] M. Gherrity, "A learning algorithm for analog, fully recurrent neural networks," in *Int. Joint Conf. Neural Networks*, 1989. IJCNN., doi: 10.1109/IJCNN.1989.118645
[82] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541-551, 1989. doi: 10.1162/neco.1989.1.4.541
[83] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning appl. to document recognition," *Proc. of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.
[84] P. Baldi, and Y. Chauvin, "Neural networks for fingerprint recognition," *Neural Computation*, vol. 5, no. 3, pp. 402-418, 1993.
[85] B. Karlik and C. Bayar, "Differentiating type of muscle movement via AR modeling and neural network classification," Turkish J. of Elect. Eng. & Computer Sci., vol. 7, no. 1-3, pp. 45-52, 1999.
[86] W. K. Chiang, D. Zhang, and L. Zhou, "Predicting and explaining patronage behavior toward web and traditional stores using neural networks: a comparative analysis with logistic regression," *Decision Support Syst.*, vol. 41, no. 2, pp. 514-531, 2006. doi: https://doi.org/10.1016/j.dss.2004.08.016
[87] M. Behrman, R. Linder, A. H. Assadi, B. R. Stacey, and M.-M. Backonja, "Classification of patients with pain based on neuropathic pain symptoms: Comparison of an artificial neural network against an established scoring system," *Eur. J. of Pain*, vol. 11, no. 4, pp. 370-376, 2007. doi: https://doi.org/10.1016/j.ejpain.2006.03.001
[88] D. E. Williams, G. E. Hinton, and R. J. Williams "Learning representations by back-propagating errors," Nature, vol. 323, no. 9, pp. 533-536, 1986.
[89] J. Hadamard, "Mémoire sur le problème d'analyse relatives à l'équilibre des plaques élastiques encastrées: Imprimerie Nationale," 1908.
[90] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko, " The math. theory of optimal processes - Int. series of monographs in pure and appl. mathematics,", New York, USA: Interscience, 1962.
[91] S. Director, and R. Rohrer, "Automated network design-the frequency-domain case," *IEEE Trans. on Circuit Theory*, vol. 16, no. 3, pp. 330-337, 1969. doi: 10.1109/TCT.1969.1082967
[92] H. J. Kelley, "Gradient theory of optimal flight paths," Ars J., vol. 30, no. 10, pp. 947-954, 1960. doi: https://doi.org/10.2514/8.5282
[93] S. Dreyfus, "The computational solution of optimal

control problems with time lag," IEEE Transactions on *Autom. Control*, vol. 18, no. 4, pp. 383-385, 1973. doi: 10.1109/TAC.1973.1100330
[94] B. Speelpenning, "Compiling fast partial derivatives of functions given by algorithms," University of Illinois at Urbana (USA). *Dept. of Comput. Sci.*, IL, USA, 1980.
[95] D. Parker, "Learning logic: tech. rep. TR-87, Center for Computational Res. in Econ. and Manage. Sci.," The MIT Press, Cambridge (Mass), 1985.
[96] D. Fletcher, and E. Goss, "Forecasting with neural networks: an application using bankruptcy data," *Inform. & Manage.*, vol. 24, no. 3, pp. 159-167, 1993. doi: https://doi.org/10.1016/0378-7206(93)90064-Z
[97] R. J. Williams, "Complexity of exact gradient computation algorithms for recurrent neural networks Technical Report Tech. Rep. NU-CCS-89-27". Boston, Northeastern University, College of Compu. Sci., Boston, USA, 1989.
[98] B. A. Pearlmutter, "Gradient calculations for dynamic recurrent neural networks: A survey," *IEEE Trans. on Neural Networks*, vol. 6, no. 5, pp. 1212-1228, 1995.
[99] A. F. Atiya, and A. G. Carlos, "New results on recurrent network training: unifying the algorithms and accelerating convergence," *IEEE Trans. on Neural Networks*, vol. 11, no. 3, pp. 697-709, 2000. doi: 10.1109/72.846741
[100] S. Amari, "Natural gradient works efficiently in learning," in *Neural Computation*, vol. 10, no. 2, , MIT Press: CA, USA. pp. 251-276, 1998. doi: 10.1162/089976698300017746
[101] M. A. Mohammed, Z. H. Salih, N. ??pu?, & R. A. K. Hasan, "Security and accountability for sharing the data stored in the cloud," in *15th RoEduNet Conf.: Networking in Edu. and Res.*," 2016. doi: 10.1109/RoEduNet.2016.7753201
[102] M. A. Mohammed, N. Ţăpuş, & R. A. K. Hasan, A. H. Omer, "A comprehensive study: Ant Colony Optimization (ACO) for Facility Layout Problem," in *16th RoEduNet Conf.: Networking in Edu. and Res.*, 2017.

About the Authors

Muneef Abdullah Ahmed is a Research Scholar with the University Polytechnic of Bucharest ,Romania, BA: University of Technology, Baghdad-Iraq.

He completed Computer Engineering in 1992 from Master University of technology Baghdad-Iraq and Computer Engineering - Software in 2005.



Professor Stefan Trăuşan-Matu is a Senior Researcher and Professor at University Politehnica of Bucharest.

He is a 2005 Fulbright Scholar. He received his post-doctoral degree from Drexel University, Philadelphia, USA. He completed Ph.D. in Computer Science in 1994 from the Department of Computer Science, "POLITECHNICA" University of Bucharest, Romania. The subject was Intelligent Reverse Software Engineering.

He completed Engineer in Computer Science from Polytechnic Institute of Bucharest in 1983.