Deep Convolutional Neural Networks versus Multilayer Perceptron for Financial Prediction

Victor-Emil Neagoe, Adrian-Dumitru Ciotec, and George-Sorin Cucu
Department of Applied Electronics and Information Engineering
Polytechnic University of Bucharest
Bucharest, Romania
Contact author email: victoremil@gmail.com

***Abstract*—This paper presents a new approach to apply and evaluate Deep Learning (DL) Convolutional Neural Networks (CNN) versus Multilayer Perceptron (MLP) for financial prediction. We have designed and evaluated a credit scoring model based on neural network classifiers in two variants: (a) MLP with eight layers; (b) DCNN with thirteen layers (six main layers and seven secondary layers). The experiments have used the German credit dataset and the Australian credit dataset. The model performances are evaluated by the following indices: Overall Accuracy (OA); False Alarm Rate (FAR); Missed Alarm Rate (MAR). The experimental results have confirmed the effectiveness of the proposed approach, pointing out the significant advantage of DCNN over the MLP. For German credit dataset, the DCNN leads to the best OA of 90.85%, versus the corresponding best MLP performance of only 81.20%. For Australian credit dataset, the DCNN has led to the best OA of 99.74%, while the MLP has obtained the best corresponding OA of 90.75%.**

Keywords—financial prediction, credit scoring, Multilayer Perceptron (MLP); Deep Learning (DL); Convolutional Neural Networks (CNN)

I. Introduction

The banks are aware of various risks [1], [2], [3]. As a consequence, the bank has to analyze the corresponding risk factors in order to optimize its decisions. The capacity to predict business failure is crucial, since incorrect decisions can lead to direct financial consequences. Credit scoring focusing on credit admission evaluation represents a serious task for financial institutions [1], [2], [3], [4]. A credit scoring model has as aim to decide whether to grant a credit to a client, taking into account the customer's features, such as income, age, marital status, education, employment status, number of existing credits, and so on.

The introduction of modern technologies has made significant changes in bank business [2]. Machine learning and data mining classifiers are used with success for financial models [1], [2], [3], [4], [5], [6], [7], [8], [9]. The stage of learning (training) consists of computation of the model parameters that approximate the mapping between input-output examples given by the labeled training set. After model learning, it can classify an unknown input sample [8], [9]. A multilayer perceptron (MLP) is a feedforward artificial neural network containing at least three layers of neurons (input layer, hidden layer, and output layer). Except for the input neurons, each neural unit uses a nonlinear activation function. MLP utilizes a supervised learning technique called backpropagation for training [8], [9], [10], [11]. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can recognize data that is not linearly separable [8], [9], [10], [11].

The solution of many artificial intelligence (AI) classification tasks depends on designing an appropriate set of features to be extracted [12]. One solution for this difficult problem is to use machine learning not only to solve the mapping from input representation to output but to discover the representation itself. This approach is named representation learning [12]. It allows AI systems to fast adaptation for new tasks with a minimum of human intervention. Representation learning leads to much better performance; it can faster discover a good set of features from raw data. Deep learning (DL) solves the problem of representation learning by the introduction of the representations that are expressed in terms of other, simpler representations [12], [13]. DL is a kind of machine learning that improves the computer with experience and data [12]. From some point of view, a DL architecture can be considered as a deeper extension of the MLP. Depth allows the computer to learn a multi-step program. Convolutional Neural Networks (CNN) are a special kind of DL feed-forward neural networks characterized by a grid-like topology and by the property of using convolution in at least one of their layers [12], [13]. A typical CNN involves four types of layers: convolutional, activation, pooling and fully-connected (dense) layers.

Within this paper we propose a neural network classifier model for credit scoring in two variants. First method uses a MLP with eight layers (input layer; six hidden layers; output layer). Second method is based on the DCNN architecture with six main layers: (input layer, three long short-term memory layers, and two fully connected layers) and other seven secondary layers (four dropout layers, Relu layer, softmax layer, and classification layer). For experimental evaluation of the proposed model, we have used two publicly downloadable datasets: German credit dataset (1000 two-class data in numerical form with 25 features each, including class label) [14], and also the Australian credit card assessment dataset (690 two-class data with 15 features each, including binary class label) [15].

978-1-5386-2350-3/18/$31.00 ©2018 IEEE

II. Proposed Neural Network Classifier Model for Credit Scoring

We further present two variant of the proposed credit scoring model, using two class supervised classification of customer feature vectors corresponding to the labels “good” or “bad”.

A. Multilayer Perceptron (MLP) Classifier for Credit Scoring

We have chosen an architecture of maximum eight layers MLP:

– input layer with ***m*** neurons, where m is the number of customer’s features

– H hidden layers with ***NH*** neurons each of them (max(H)=6)

– output layer with two neurons (corresponding to the decisions “good” and “bad”).

We propose to perform MLP training with Broyden-Fletcher-Goldfarb-Shannon (BFGS) quasi-Newton algorithm [16], that has proved to lead to significantly better performances than those obtained by classical Levenberg-Marquardt backpropagation [11].

B. Deep Convolutional Neural Networks (DCNN) Classifier for Credit Scoring

We propose a credit scoring classifier architecture composed by the following thirteen DCNN layers:

• *input* layer with ***m*** neurons (corresponding to the number of client’s features);

• *three main layers* of the kind *Long Short-Term Memory (LSTM),* from layer 2 to layer 4, with ***N*** neurons each of them, each of the above main layers being associated with a corresponding secondary *dropout layer* with the probability D.

• Fully Connected (FC) layer (main layer number 5) with N neurons, the FC layer being associated with two secondary layers:

– *Dropout layer* with the probability D

– *Relu layer*

• Fully Connected (FC) layer (main layer number 6) with two neurons, the FC layer being associated with two secondary layers:

– *Softmax layer*

– *Classification layer*

Inclusion of the dropout layers is a consequence of the regularization approach used to reduce the overfitting.

We have chosen the Stochastic Gradient Descent with Momentum for DCNN training.

III. Experiments and Results

A. Experimental Setup

We have evaluated the credit scoring performances according to the proposed model in terms of the indicators defined below {Overall Accuracy (OA), Miss Alarm Rate (MAR) and False Alarm Rate (FAR)}

Denote:

– TP (True Positives) = number of test data correctly classified as “bad”;

– TN (True Negatives) = number of test data correctly classified as “good”;

– FP (False Positives) = number of test data incorrect classified as “bad”.

– FN (False Negatives) = number of test data incorrect classified as “good”.

– NT=TP+TN+FP+FN=total number of test data

Define:

– False Alarm Rate (FAR) (%)

$$FAR = FP/(FP+TN)\times 100\ [\%] \quad (1)$$

– Miss Alarm Rate (MAR) (%):

$$MAR = FN/(FN+TP)\times 100\ [\%] \quad (2)$$

– Overall Accuracy (OA) (%):

$$OA = (TP+TN)/\ N_T \times 100\ [\%] \quad (3)$$

B. German Credit Dataset

•Dataset Description

We have firstly considered the German credit dataset provided by Strathclyde University in the variant of numerical attributes, included in the file “German.data-numeric” [14]. The considered dataset (numerical variant) is composed by 1000 vectors (data) with 25 features each, including customer label (1=good; 2=bad). The other **m=24** customer’s features correspond to the numerical encoding of the characteristics as: age, income, employment, marital status and sex, properties (house, car), qualifications (skills); existing credits, and so on.

•Experimental Results

The experimental results for German dataset are given in Tables I, II and III. We have chosen the OA performance (yellow column) as a main target.

In Table I, one can evaluate MLP credit scoring performances. One deduces that by increasing the number of hidden layers, the OA is improved. The best OA performance of 81.20 % corresponds to the maximum of H=6 hidden layers. One can also remark, that the best performance (for a given number of H layers) corresponds to choose the number of hidden neurons in the interval [12, 18].

Table II shows that choosing a MLP architecture with all the six hidden layers, the OA performance does not usually depend too much on the number of neurons NH characterizing each hidden layer.

Table III shows better performances of DCNN classifier for credit scoring, taking the MLP as a reference. We have marked by red color the cases with OA over 90%.

By increasing the number N of neurons of each of the main layers {2,3,4,5}, one remarks a general increasing of performances and the increasing of the number of cases with OA greater than 90%.

The best performance regarding OA is of 90.85% and this corresponds to choose a number of N=296 neurons for each layer.

By considering only the cases with OA higher than 90%, the best MAR (lower is better) is of 3.81%, and it corresponds to N=362.

TABLE I. BEST MLP CREDIT SCORING OVERALL ACCURACY (OA) AS A FUNCTION OF THE NUMBER OF HIDDEN LAYERS (GERMAN CREDIT DATASET) (MLP TRAINING WITH BFGS QUASI-NEWTON ALGORITHM; **NL=500** TRAINING DATA; **NT=250** TEST DATA)

Number of hidden layers (H)	Number of neurons of any hidden layer for best OA (NH)	Overall Accuracy (OA) [%]	False Alarm (FAR) [%]	Missed Alarm (MAR) [%]
1	16	78.40	14.12	39.73
2	16	76.80	18.46	40
3	12	77.60	17.13	36.23
4	14	78.40	19.81	30.23
5	18	77.60	18.96	41.03
6	**18**	**81.20**	13.98	32.81

TABLE II. MLP CREDIT SCORING OVERALL ACCURACY (OA) AS A FUNCTION OF THE NUMBER OF NEURONS FOR EACH OF THE HIDDEN LAYERS (GERMAN CREDIT DATASET; MLP TRAINING WITH BFGS QUASI-NEWTON ALGORITHM; **NL=500** TRAINING DATA**; NV= 250** VALIDATION DATA; **NT**=250 TEST DATA; H=6 HIDDEN LAYERS)

Number of neurons of any hidden layer for best OA (NH)	Overall accuracy (OA) for training set [%]	Overall accuracy (OA) for validation set [%]	Overall Accuracy (OA) for test set [%]	Overall Accuracy (OA) for all data [%]
1	73.20	70.40	68.80	71.40
2	72.80	67,20	70.80	70.90
4	77.00	72.80	72.00	74.70
6	76.20	72.40	73.20	74.50
8	79.40	71.20	77.60	76.90
10	76.60	70.80	74.80	74.70
12	81.40	72.80	72.00	76.90
14	82.40	70.80	72,.40	7.00
16	76.00	75.60	71.20	74.70
18	78.40	75.20	**81.20**	78.30
20	79.20	75.60	70.80	76.20
22	79.20	75.60	70.80	76.20

TABLE III. DCNN CREDIT SCORING PERFORMANCES AS A FUNCTION OF THE NUMBER OF NEURONS FOR EACH OF THE MAIN LAYERS {2,3,4,5} (GERMAN CREDIT DATASET; DCNN TRAINING USING THE ALGORITHM OF STOCHASTIC GRADIENT DESCENT WITH MOMENTUM; **NL=300** TRAINING DATA; **NT**=700 TEST DATA)

Number of neurons N of each of layers {2,3,4,5}	Overall Accuracy (OA) [%]	False Alarm (FAR) [%]	Missed Alarm (MAR) [%]
130	86.42	30.00	9.00
142	83.42	53.33	6.54
155	87.14	33.33	7.27
158	**90.00**	**18.66**	**7.63**
162	86.85	20.00	11.27
167	85.28	25.33	11.81
177	84.28	8.66	17.63
181	87.42	9.33	13.45
190	88.42	28.66	6.90
194	85.57	16.66	13.81
203	87.71	32.66	6.72
204	88.14	20	9.63
212	88.42	21.33	8.90
217	87.42	18.00	11.09
223	88.42	18.66	9.63
234	87.28	17.33	11.45
237	**90.57**	**16.66**	**7.45**
250	89.42	22.00	7.45
257	89.42	17.33	8.72
284	87.14	2.00	15.81
286	**90.28**	**17.33**	**7.63**
293	**88.28**	**23.33**	**8.54**
296	**90.85**	**32.00**	**2.90**
304	89.14	5.33	12.36
313	89.42	5.33	12.00
320	**90.00**	**5.333**	**11.27**
328	89.14	20.66	7.45
331	88.28	2.66	14.18
348	**90.14**	**20.66**	**6.90**
357	87.57	8.667	13.45
361	**90.14**	**19.33**	**7.27**
362	**90.14**	**32.00**	**3.81**
376	89.42	34.00	4.18
379	**90.42**	**18.66**	**7.09**
382	89.85	18.66	7.81
387	**90.14**	**16.66**	**8.00**
392	87.57	7.33	13.81
394	**90.14**	**20.00**	**7.09**
396	89.57	15.33	9.09
399	**90.28**	**18.66**	**7.27**

C. Australian Credit Dataset

•Dataset Description

We have also considered the Australian credit card assessment dataset [15], containing 690 patterns (vectors) with **m=14** attributes; 6 numeric and 8 discrete (with 2 to 14 possible values). The 15th feature is the binary class label.

•Experimental Results

The experimental results for Australian credit approval dataset are given in Tables IV and V.

Table IV shows that the OA performance of Australian credit dataset for a MLP with eight layers is not influenced too much by the number of hidden neurons. The best OA of 90.75 % is obtained for NH=41 neurons. By considering only cases with OA bigger than 89%, one can deduce that the best (minimum) MAR is of 8.64% and it is obtained by choosing NH=22 neurons.

Table V shows the very good performances of DCNN classifier for credit scoring, significantly higher than those obtained by MLP. We have marked by red color the cases with OA over 99%.

By choosing the number N of neurons of the main layers {2,3,4,5} in the interval [65, 129], we have obtained a stable behavior of the proposed DCNN architecture, with a significant number of cases with OA higher than 99%. The best OA of 99.74% is obtained for N=85 and N=98; for the same two higher performance cases, one obtains MAR of 0% and FAR of 0.63%.

TABLE IV. MLP CREDIT SCORING PERFORMANCES FOR THE AUSTRALIAN CREDIT APPROVAL **TEST** DATASET AS A FUNCTION OF THE NUMBER OF NEURONS FOR EACH OF THE HIDDEN LAYERS (MLP TRAINING WITH BFGS QUASI-NEWTON ALGORITHM; **NL=344** TRAINING DATA**; NT=173** TEST DATA; H=6 HIDDEN LAYERS)

Number of neurons of any hidden layer for best OA (NH)	Overall Accuracy (OA) [%]	False Alarm (FAR) [%]	Missed Alarm (MAR) [%]
1	87.86	12.12	12.16
2	88.44	2.63	18.56
3	83.82	14.00	19.18
4	86.13	11.88	16.67
5	**89.02**	4.65	17.24
6	87.86	8,70	16.05
7	87.86	16.33	6.67
8	81.50	11.70	26.58
9	87.28	13.13	12.16
10	83.82	12.24	21.33
11	88.44	8.64	14.13
12	86.13	11.43	17.65
13	87.86	10.99	13.41
14	77.46	20.51	24.21
15	87.28	9.09	16.47
16	83.24	10.42	24.68
17	83.24	20,00	12.82
18	87.28	7.61	18.52
19	86.13	10.59	17.05
20	81.50	19.79	16.88
21	86.13	16.83	9.72
22	**89.60**	**11.96**	**8.64**
23	88.44	14.13	8.64
24	**89.02**	**7.14**	**16.00**
25	79.77	15.24	27.94
26	84.39	15.91	15.29
27	84.97	16.48	13.41
28	87.86	12.35	11.96
29	87.86	10.99	13.41
30	84.39	13.48	17.86
31	84.39	13.33	18.07
32	84.97	11.32	20.90
33	83.24	12.50	22.08
34	**89.02**	**10.47**	**11.49**
35	81.50	11.83	26.25
36	86.71	10.47	16.09
37	84.97	16.50	12.86
38	82.66	9.20	25.58
39	83.82	13.73	19.72
40	83.24	9.78	24.69
41	**90.75**	**8.91**	**9.72**
42	87.86	8.70	16.05
43	83.82	11.90	20.22
44	85.55	10.59	18.18
45	87.28	12.38	13.24

TABLE V. DCNN CREDIT SCORING PERFORMANCES AS A FUNCTION OF THE NUMBER OF NEURONS FOR EACH OF THE LAYERS {2,3,4,5} (AUSTRALIAN CREDIT APPROVAL DATASET; DCNN TRAINING USING THE ALGORITHM OF STOCHASTIC GRADIENT DESCENT WITH MOMENTUM; **NL=300** TRAINING DATA (150 CREDIT APPROVAL +150 CREDIT REJECTION) ; **NT=390** TEST DATA)

Number of neurons N of each of main layers {2,3,4,5}	Overall Accuracy (OA) [%]	False Alarm (FAR) [%)]	Missed Alarm (MAR) [%]
64	95.64	0	7.29
65	**99.23**	**1.91**	**0**
66	92.82	0	12.01
67	98.72	3.18	0
68	95.64	3.18	5.15
69	96.41	0	6.00
70	**99.23**	**1.91**	**0**
71	98.97	2.54	0
72	93.33	16.56	0
73	94.36	0	9.44
74	92.82	0	12.01
75	93.59	0	10.72
76	**99.23**	**0.63**	**0.85**
77	**99.49**	**0.63**	**0.42**
78	**99.23**	**1.91**	**0**
79	**99.49**	**0.63**	**0.42**
80	**99.49**	**1.27**	**0**
81	**99.23**	**1.91**	**0**
82	98.72	**0**	**2.14**
83	**99.23**	**1.91**	**0**
84	**99.23**	**1.27**	**0.42**
85	**99.74**	**0.63**	**0**
86	98.72	0	2.14
87	98.72	1.27	1.28
88	**99.23**	1.91	0
89	98.97	1.91	0.42
90	98.72	3.18	0
91	**99.23**	**0.63**	**0.85**
92	**99.23**	**1.91**	**0**
93	**99.23**	**1.27**	**0.42**
94	**99.23**	**1.91**	**0**
95	**99.23**	**1.91**	**0**
96	98.97	2,54	0
97	95.13	0	8.15
98	**99.74**	**0.63**	**0**
99	97.69	5.09	0.42
100	98.97	2.54	0

IV. CONCLUDING REMARKS

This paper proposes a new neural network classifier model for financial prediction in two variants. First variant uses a MLP with eight layers trained with Broyden-Fletcher-Goldfarb-Shannon (BFGS) quasi-Newton algorithm. Second variant implies a DCNN architecture with thirteen layers (six main layers and seven secondary layers), using a Stochastic Gradient Descent with Momentum algorithm for training.

The experimental results have confirmed the effectiveness of the proposed approach. The performance difference between the two variants is significant; one can clearly point out the important advantage of DCNN over the MLP.

For the German credit dataset, the DCNN leads to the best OA of 90.85% versus the corresponding best MLP performance of only 81.20%. The DCNN has led to the best (minimum) MAR of 3.81% (having at the same time OA greater than 90%), while the MLP has obtained a minimum MAR of only 32.81%. On the other side, the best overall classification scores of over 90% obtained by our credit scoring prediction model based on DCNN are clearly higher than the average classification score of 85.33% reported in literature for German credit data [1].

For the Australian credit dataset, the proposed DCNN leads to the best OA of 99.74%, in comparison with the corresponding best MLP performance of 90.75%. The DCNN has led to the best (minimum) MAR of 0% versus the best MAR obtained by MLP of 8.64%. At the same time, the best overall classification score of 99.74 % obtained by the proposed credit scoring prediction model based on DCNN is significantly higher than the average classification score of 89.59% reported in literature for Australian credit data [1].

REFERENCES

[1] W.Y. Lin, Y. H. Hu, C. F. Tsai, "Machine Learning in Financial Crisis Prediction: A Survey," IEEE Trans. Systems, Man, and Cybernetics– Part C: Applications and Reviews, Vol. 42, No. 4, July pp. 421-436, July 2012.

[2] V. Jayasree, R. V. S. Balan, "A review on data mining in banking sector," American Journal of Applied Sciences vol. 10 , no 10, pp. 1160-1165, 2013.

[3] T. Damrongsakmethee, V.E. Neagoe, "Data Mining and Machine Learning for financial analysis," Indian Journal of Science and Technology, Vol 10(39), pp. 1-7, October, 2017.

[4] N.C. Hsieh, "Hybrid mining approach in the design of credit scoring models," Expert Systems with Applications, Vol. 28, pp. 655–665, 2005.

[5] V. E. Neagoe, R. Iatan, I.F. Iatan, "A nonlinear neuro-fuzzy model for prediction of daily exchange rates," Proc. 5th International Symposium on Soft Computing for Industry, held at the World Automation Congress (WAC'04), Seville, June 28-July 1, 2004, published in : Soft Computing with Industrial Applications, Vol. 17, Book Series: TSI Press, pp. 573-578.

[6] L. Di Persio, and O. Honchar. "Artificial neural networks architectures for stock price prediction: Comparisons and applications." International Journal of Circuits, Systems and Signal Processing, Vol. 10, pp. 403-413, 2016.

[7] G. Suciu, C. Dobre, V. Suciu, G.Todoran, A. Vulpe, A. Apostu. "Cloud computing for extracting price knowledge from big data," Proc. Ninth International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS), Blumenau, Brazil, July 8-10, 2015, pp. 314-317.

[8] M. Bishop, Pattern Recognition and Machine Learning, New York: Springer, 2006.

[9] S. Shalev-Shwartz, S. Ben-David, Understanding Machine Learning-From Theory to Algorithms, New York: Cambridge Press, 2014.

[10] D.E. Rumelhart, G.E. Hinton, R. J. Williams. "Learning internal representations by error propagation", D. E. Rumelhart, J.L. McClelland, and the PDP research group. (editors), Parallel distributed processing: Explorations in the microstructure of cognition, Vol.1: Foundation, Cambridge, MA, USA: MIT Press, 1986.

[11] S. Lahmiri, "A comparative study of backpropagation algorithms in financial prediction," International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.1, No.4, pp. 15-21, August 2011.

[12] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, Cambridge, MA, USA: MIT Press, 2016.

[13] J. Patterson, A. Gibson, Deep Learning-A Practitioner's Approach, Sebastopol, California: O'Reilly Media, Inc., 2017.

[14] http://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data)

[15] http://archive.ics.uci.edu/ml/datasets/Statlog+(Australian+Credit+Approval)

[16] N. N. Schraudolph, J. Yu, S. Günter, "A Stochastic Quasi-Newton Method for Online Convex Optimization," Proc. 11th International Conf. Artificial Intelligence and Statistics (AISTATS), pp. 433–440, San Juan, Puerto Rico, March 2007.