Proceedings of CONF-MLA 2026 Symposium: Intelligent Systems and Automation: AI, IoT, Robotic Engineering & Algorithm
DOI: 10.54254/2755-2721/2026.LD35067

An Analysis of the Practicality of DNN Models versus LR Models in Credit Scoring

Shenchen Fei

Faculty of Science and Engineering, University of Nottingham Ningbo China, Ningbo, China
feisc2004@outlook.com

Abstract. Financial technology is playing an increasingly vital role in loan decision-making, and financial institutions are increasingly relying on machine learning techniques to support credit decisions. The purpose of this review is to provide a critical overview of analysis comparing the practical applicability of deep neural network (DNN) and logistic regression (LR) models within the credit scoring domain. This paper systematically collects existing studies on the application of DNN and LR models in credit scoring. The research objects include DNN and LR models, as well as their improved variants developed on the original model frameworks. On this basis, it integrates theoretical research findings with comprehensive analyses to investigate and evaluate the practicality of DNN models. The research results indicate that DNN models still exhibit significant limitations in credit scoring applications. Further model improvements or hybrid integration with other models are therefore required to enhance their practical applicability in real-world scenarios.

Keywords: credit scoring, logistic regression, deep neural network

1. Introduction

The importance of financial technology in decision-making is becoming increasingly evident, and financial institutions are increasingly relying on machine learning to inform their decisions [1, 2]. In this area, the statistical credit scoring method was first proposed and applied by Durand. As core technical tools in the credit scoring domain, machine learning models have undergone continuous iteration, including logistic regression, decision trees, and neural networks, which have profoundly promoted the long-term development of credit risk assessment [3]. Nevertheless, although state-of-the-art deep neural network (DNN) architectures have been proposed for quite some time, real application cases of DNN models remain far fewer in number compared with logistic regression (LR) models. LR models still occupy a dominant position in the field of credit scoring. Accordingly, this paper primarily focuses on a comparative analysis of LR and DNN models. It aims to explore the underlying causes of the limited practical adoption of DNN models and provide feasible insights and optimization suggestions for improving their practical applicability in credit scoring. By investigating the superiorities of logistic regression variants and analyzing the inherent limitations of conventional DNN models, this paper elaborates on the core reasons for the low practical adoption of DNN models in credit scoring. Furthermore, combined with advanced DNN variants, this study

DOI: 10.54254/2755-2721/2026.LD35067

offers a general overview of the current development status, and provides feasible prospects and developmental directions for the future application of DNN-based models in this field.

2. Practical background and model evolution of credit scoring

Gunnarsson et al. elaborated on the practical context of credit scoring in the research. At present, commercial banks possess plenty of consumer loans, rendering consumer credit an industry of great economic significance. According to the article of Tu and Wu, data released by the Federal Reserve shows that U.S. consumer credit exceeded $4.7857 billion in 2022, maintaining a sustained annual growth rate of over 4% [4]. Thus, credit scoring acts as a key determinant of economic returns within this sector and improvements in scoring accuracy can generate tangible economic benefits. From another perspective, data released by the European Central Bank showed that fraudulent transactions with cards issued within the Single Euro Payments Area and processed globally reached a staggering 1.8 billion euro. Hence, robust credit scoring models are urgently required to prevent, manage and mitigate financial losses [4]. Against this backdrop, developing precise credit scoring models has become a top priority in this field [3]. Xiao J. et al. also stated that credit scoring serves as an effective tool for addressing credit risk management. It constitutes a binary classification problem, where loan applicants are divided into good and bad credit groups based on features, such as loan amount, occupation and income [5]. Since the emergence of credit scoring systems, numerous classification methodologies have been developed and deployed for credit risk assessment. These approaches cover classical statistical models represented by logistic regression, and machine learning frameworks such as decision trees, as well as neural network architectures. Extensive scholarly efforts over recent decades have focused on evaluating and comparing the predicative performance of different classification algorithms within credit scoring scenarios [6]. Logistic regression and deep neural networks represent two landmark architectures in the evolution of credit scoring models.

3. Fundamental introduction and comparison to LR and DNN models

The LR model is among the most frequently utilized models within the domain of the credit scoring system. Due to strong interpretability, the LR model could calculate the marginal effects with high efficiency and enable scorecards, elasticities and significance testing. In addition, LR has excellent performance on simplicity and stability, which meets regulatory requirements for model transparency [7]. It causes substantial impacts on the development of various fields, especially for finance, and facilitates economic efficiency and contributes to economic vitality at the same time. Apart from the advantages mentioned above, the LR model delivers distinct advantages in small sample settings over neural network model approaches. Bensic et al. compared the performance of LR with NN and CART decision trees' and found that the research showed an indistinct difference in proportion and McNemar's test based on 160 samples in the credit scoring context of small enterprises in Croatia [8]. This finding reveals that insufficient sample sizes prevent neural networks from exerting their practical performance advantages, in spite of their theoretically strong capability for nonlinear fitting. By contrast, LR models demonstrate greater stability and fewer constraints when applied to small-sample datasets. Though the LR model is the industry standard in credit scoring, it falls behind the demands because of the complexity of the application situations and changing data. Since the credit data often contains threshold and interaction effects, the LR model is incapable of capturing and dealing with non-linear relationships effectively. Though with quadratic or interaction terms, it requires manual features that cause inefficiency and overfitting problems [7].

DOI: 10.54254/2755-2721/2026.LD35067

It reveals that the LR model struggles with intricate non-linear structures. Therefore, a new data processing model, called DNN, proven and applied to the credit scoring area.

To adapt to data distribution diversity, the credit scoring model has developed from traditional statistical methods and traditional machine learning to DNN. The algorithms assign the DNN model the capacity of presenting features by automatic learning, which means there is no need to construct synthetic data and the model could learn useful output on its own. Secondly, capturing complex non-linear relationships between variables is another remarkable achievement. The advancement of related technologies has enabled the DNN credit scoring model to overcome the weakness of traditional LR models in handling non-linear relationships. Otherwise, the entire process could be jointly optimized from the original input to the ultimate prediction [9]. Although the DNN model has largely tackled the drawbacks of the LR model, it still faces various challenges and difficulties. The research of Mestiri indicates that the Area Under Curve (AUC) of DNN on the dataset of 688 samples reaches 0.788, which is significantly better than the LR model's performance with a value of 0.53 [6]. However, the quantity of samples is quite lower than normal research. And the model exhibits an apparent risk of overfitting. Hence, the potential risks and drawbacks of the DNN model must not be overlooked, despite its notable advantages.

4. Analysis of factors hindering practical application of DNN models

Conventional wisdom holds that when a novel technology is introduced and its advantages are broadly acknowledged, it tends to displace established alternatives and emerge as the dominant solution within its domain. However, in the field of credit scoring, the practical deployment of DNN models has not yielded performance improvements commensurate with widespread public expectations. This discrepancy arises primarily from two correlated factors: first, established variants of LR models continue to satisfy prevailing regulatory, interpretability, and performance requirements in credit scoring; second, DNN models encounter substantial practical constraints, including difficult-to-design structure, black-box nature, data imbalance and high training cost [5].

4.1. Expansions and impacts of LR variants

Although the applications of the LR model have existed for a long time and lack the ability to deal with complex non-linear relationship problems, various extended forms and modified variants of the LR model provide a suite of powerful competing models for credit scoring tasks. Beyond the binary dichotomy between the LR and DNN models, parts of advanced studies have explored methodological improvements to the LR model algorithm itself. Focusing on the prevalent linguistic variables in the credit scoring of technological enterprises, Sohn et al. gave a new LR model called the fuzzy logistic regression model. This model converts Likert scale scores into fuzzy triangular numbers and adapts fuzzy least squares estimation of coefficients. Under the precondition of saving the interpretability of LR coefficients, the fuzzy LR model improves the predictive performance significantly [10]. This research shows that LR are not basic models when facing the challenge DNN brings. Furthermore, the potential for methodological optimization deserves in-depth attention. Besides, Dumitrescu et al. proposed another available LR-based model called Penalized Logistic Tree Regression (PLTR). This model first uses shallow decision trees to automatically identify the non-linear patterns, such as specific thresholds or combinations of borrower attributes, which are converted into simple yes or no rules. Then, a penalized logistic regression selects and weighs only the most predictive rules, producing a final model that is both more accurate than LR and easier to interpret than a traditional scorecard. Simulation experiments and economic evaluations were further

DOI: 10.54254/2755-2721/2026.LD35067

conducted. The results show that the PLTR model achieves predictive performance comparable or marginally better than random forest, with superior interpretability. Meanwhile, it incurs lower costs relative to the conventional LR model [7]. This integrated model combines the merits of random forest and logistic regression. It inherits the high predictive accuracy of random forest while retaining the strong interpretability of logistic regression, and achieves an enhanced capacity to capture nonlinear effects. In summary, variants and extended frameworks of the LR model can well accommodate the current demands of credit scoring research and mitigate the drawbacks inherent to the original LR model, rendering the adoption of DNN models unnecessary.

4.2. Inherent limitations of DNN models

Aside from the existence of competitive LR variants, the inherent limitations of DNN models in credit scoring hinder their practical deployment. J. Xiao et al. mentioned four prominent limitations of DNN. These include difficult architecture design, poor interpretability stemming from its black-box nature, insufficiency for imbalanced data handling, and high training costs. Hyperparameters of DNN, such as the number of layers, activation functions and neurons per layer, need to be defined before training. However, these parameters can hardly be predicted in advance and require a number of trials and errors. Consequently, the model design workflow becomes tedious and requires experience. In the field of financial risk control, model interpretability is particularly critical, as financial institutions must clarify the underlying criteria and rationale for credit evaluation results. Nevertheless, DNN models involve highly complex internal decision processes that are difficult to comprehend. Such inherent black-box characteristics substantially limit the practical applicability of DNN models in financial scenarios. In addition, credit scoring datasets inherently suffer from severe class imbalance, where default samples occupy only a tiny proportion. The DNN model possesses no native mechanisms tailored to mitigate such imbalanced distribution. And existing research still adapts conventional resampling methods to solve the problems without proving DNN models themselves. Besides, DNNs demand huge computational resources and lengthy training durations, meanwhile, exhibiting high sensitivity to hyperparameter tuning. Since disparate datasets often call for completely unique hyperparameter setups, such models deliver poor transfer performance across data sources [5]. Venusiana et al. also addressed the black-box issue of DNNs, pointing out that DNNs are prone to overfitting, impose heavy computational requirements, and carry intrinsic black-box attributes. These drawbacks hinder compliance with financial regulatory requirements, alongside their inadequate learning capacity for class-imbalanced data [11]. Both studies identify largely identical limitations of DNN models, which partly explains the reason why DNN models have not achieved widespread practical adoption, despite theoretical advantages over LR models.

5. Optimization approaches of DNN models

The superiority of extended and variant LR models has been well established in existing research. Similarly, further expansion and optimization can be implemented on the basis of DNN models. Such improved strategies can preserve the unique nonlinear fitting advantages of DNN models while effectively mitigating the risks caused by their inherent limitations, thereby enhancing the practical applicability and providing feasible solutions for the real-world deployment of DNN-based credit scoring models. L. Wang et al. identified the weak interpretability and inherent black-box nature of DNN models, and accordingly, developed an innovative model grounded on DNN models. Hence, a novel model called an attentive contrast deep neural network (ACNet) is invented, which combines the advantages of neural network and tree-based models. ACNet is a deep neural network model

DOI: 10.54254/2755-2721/2026.LD35067

equipped with a hierarchical decision structure. With the integration of sparse feature selection and self-attention mechanisms, it dynamically picks out salient features and realizes progressive representation learning. In addition, a special credit contrastive loss function (CCL) is introduced. By pulling samples of the same category closer and pushing heterogeneous samples farther apart, the model achieves enhanced separability between different samples. Most importantly, ACNet can derive both local and global interpretability through its internal learning process, which differs from conventional DNN models that rely on external tools to generate model explanations. By adopting the CCL, ACNet effectively separates sample distributions, alleviates adverse impacts caused by imbalanced credit datasets, and simultaneously addresses the poor interpretability issue of standard DNN models [9]. Thus, it can be seen that the inherent limitations of DNN models are not insurmountable, and models can be revitalized in the analogous logic as LR variants. ACNet offers a viable blueprint for upgrading and extending DNN architectures, and it also represents a mainstream direction for the practical deployment of DNN-based models in credit scoring research moving forward.

6. Conclusion

This article has synthesized the existing literature on the application of LR and DNN models in the credit scoring area, with a main focus on a comprehensive comparison of the strengths and weaknesses of the LR and DNN models and further exploration of the underlying reasons for the practical application of DNN models in this area. The existing evidence yields two vital conclusions: (1) the LR model is still the standard model utilized in credit scoring; (2) DNN models suffer from notable inherent limitations. However, there remains a divergence of opinions regarding the applicability of DNN models in current credit scoring practices. Several studies have demonstrated that improved and variant LR models can adequately satisfy the requirements of credit scoring, thereby reducing the necessity for adopting DNN models. Conversely, other studies have validated that optimized DNN models possess considerable application potential and can deliver superior performance for credit scoring tasks. Such discrepancies may stem from differences in practical demands, specific application scenarios, and data distribution characteristics. Nevertheless, substantial research gaps still exist in current studies. First, limited efforts have been devoted to optimizing the training cost of DNN models. Second, there is a lack of unified evaluation criteria for the performance comparison of different credit scoring models. Third, existing literature contains insufficient practical analysis of DNN model applications, which fails to intuitively demonstrate the pros and cons of DNN models in credit scoring. Future research should focus on developing diverse DNN variants to substantially mitigate their inherent limitations and establishing standardized evaluation protocols for model assessment.

References

[1] Kozodoi, N., Jacob, J., Lessmann, S. (2022) Fairness in credit scoring: Assessment, implementation and profit implications. European Journal of Operational Research, 297: 1083-1094.

[2] Gambacorta, L., Huang, Y., Qiu, H., Wang, J. (2024) How do machine learning and non-traditional data affect credit scoring? New evidence from a Chinese fintech firm. Journal of Financial Stability, 73: 101284.

[3] Gunnarsson, B.R., Broucke, S.V., Baesens, B., Óskarsdóttir, M., Lemahieu, W. (2021) Deep learning for credit scoring: Do or don't? European Journal of Operational Research, 295: 292-305.

[4] Tu, J., Wu, Z. (2025) Inherently interpretable machine learning for credit scoring: Optimal classification tree with hyperplane splits. European Journal of Operational Research, 322: 647-664.

[5] Xiao, J., Zhong, Y., Jia, Y., Wang, Y., Li, R., Jiang, X., Wang, S. (2024) A novel deep ensemble model for imbalanced credit scoring in internet finance. International Journal of Forecasting, 40: 348-372.

[6] Mestiri, S. (2024) Credit scoring using machine learning and deep Learning-Based models. Data Science in Finance and Economics, 4(2): 236-248.
[7] Dumitrescu, E., Hué, S., Hurlin, C., Tokpavi, S. (2022) Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. European Journal of Operational Research, 297: 1178-1192.
[8] Bensic, M., Sarlija, N., Zekic-Susac, M. (2005) Modelling small-business credit scoring by using logistic regression, neural networks and decision trees. Intell. Sys. Acc. Fin. Mgmt., 13: 133-150.
[9] Wang, L., Xu, S. Zhang, L. (2026) Attentive contrast deep neural network for imbalanced credit scoring. Engineering Applications of Artificial Intelligence, 163: 112917.
[10] Sohn, S.Y., Kim, D.H., Y, J.H. (2016) Technology credit scoring model with fuzzy logistic regression. Applied Soft Computing, 43: 150-158.
[11] Venusiana, F.M.R., Wibowo, A., Sofwan, A. (2026) A Hybrid Decision Tree Deep Neural Network (DT-DNN) model with Quadratic Activation Function for SME Credit Scoring. Engineering, Technology & Applied Science Research, 16(2): 33902-33908.