IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 19, 2020, accepted January 9, 2021, date of publication January 12, 2021, date of current version January 19, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3051079

Deep Learning Anti-Fraud Model for Internet Loan: Where We Are Going

WEIWEI FANG[1,3], XIN LI[2,4], PING ZHOU[2], JINGWEN YAN[2,4], DAZHI JIANG[2,4], AND TENG ZHOU[2,4]

[1]College of Science, Shantou University, Shantou 515063, China
[2]College of Engineering, Shantou University, Shantou 515063, China
[3]College of Information Engineering, Nanyang Institute of Technology, Nanyang 473000, China
[4]Key Laboratory of Intelligent Manufacturing Technology, Ministry of Education, Shantou University, Shantou 515063, China

Corresponding authors: Teng Zhou (zhouteng@stu.edu.cn) and Xin Li (lixin@stu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61902232, in part by the Science and Technology Plan Project of Henan Province under Grant 182102210459, in part by the Key Scientific Research Plan Project of Henan University under Grant 16A510007, in part by the Natural Science Foundation of Guangdong Province under Grant 2018A030313291, in part by the Education Science Planning Project of Guangdong Province under Grant 2018GXJK048, in part by the STU Scientific Research Foundation for Talents under Grant NTF18006, and in part by the 2020 Li Ka Shing Foundation Cross-Disciplinary Research under Grant 2020LKSFG05D.

ABSTRACT Recently, Internet finance is increasingly popular. However, bad debt has become a serious threat to Internet financial companies. The fraud detection models commonly used in conventional financial companies is logistic regression. Although it is interpretable, the accuracy of the logistic regression still remains to be improved. This paper takes a large public loan dataset, *e.g.* Lending club, for example, to explore the potential of applying deep neural network for fraud detection. We first fill the missing values by a random forest. Then, an XGBoost algorithm is employed to select the most discriminate features. After that, we propose to use a synthetic minority oversampling technique to deal with the sample imbalance. With the preprocessed data, we design a deep neural network for Internet loan fraud detection. Extensive experiments have been conducted to demonstrate the outperformance of the deep neural network compared with the commonly-used models. Such a simple yet effective model may brighten the application of deep learning in anti-fraud for Internet loans, which would benefit the financial engineers in small and medium Internet financial companies.

INDEX TERMS Internet finance, loan fraud detection, deep learning, financial model.

I. INTRODUCTION

Internet fraud methods are increasing dramatically in recent years, together with the rapid development of Internet financial models and the Internet business used to be handled by traditional financial institutions. In this regard, Internet lending companies face an unprecedented risk of online fraud. Luckily, the rapid development of computer technology, the accumulating data, and the emerging data analysis techniques bring new opportunities to financial risk management and analysis on the big data in the financial industry.

Researchers have developed various anti-fraud measures and fraud prevention systems over the years. Leonard [1] proposed a rule-based expert system for fraud detection. The rules of this model were manually constructed by the fraud experts from the bank. Sanchez *et al.* [2] proposed to use association rules to detect fraud and help risk analysts extract more fraud rules. Edge and Sampaio [3] proposed a set of a financial fraud modeling language (FFML) for better describing and combining fraud rule sets to assist fraud analysis. However, the rule-based models require sufficient and accurate expertise knowledge and can not be updated timely to new frauds.

To this end, machine learning models have been introduced for fraud detection. Ghosh and Reilly [4] uses neural networks to detect credit card fraud. Kokkinaki [5] proposed decision trees and Boolean logic functions to characterize normal transaction patterns to detect fraudulent transactions. Peng *et al.* [6] compared nine machine learning models for fraud detection. The results demonstrate linear logistic and

The associate editor coordinating the review of this manuscript and approving it for publication was Xiong Luo.

Bayesian networks are more effective. Lei and Ghorbani [7] proposed a new clustering algorithm namely improved competitive learning network (ICLN) and supervised an improved competitive learning network (SICLN). Sahin *et al.* [8] designed a decision tree based on cost sensitivity. Halvaiee and Akbari [9] proposed to use an AIRS improved algorithm for fraud detection. However, these traditional machine learning methods heavily rely on manual subjective rules and easily lead to model risk. These methods also tend to overfit due to the imbalance training dataset with serious pollution by noises. Thus, ensemble learning methods have also been introduced to integrate different models for complicated fraud detection. Louzada and Ara [10] proposed a bagging ensemble model that integrates k-dependence probabilistic networks. The results show that the proposed ensemble model has stronger modeling capabilities. Carminati *et al.* [11] proposed a combination of semi-supervised and unsupervised fraud and anomaly detection methods, mainly using a histogram-based outlier score (HBOS) algorithm to model the user's past behavior.

Recently, deep learning techniques have attracted a lot of academic and industrial attention that provides a new insight for financial data analysis. Fu *et al.* [12] used convolutional neural networks to effectively reduce feature redundancy. Tu *et al.* [13] design a deep feature representation technique for fraud detection. To incorporate with prior knowledge with the deep network, Greiner and Wang [14] pointed out the borrower is likely to conceal information that is not beneficial to him or even fictitious favorable information before obtaining the loan. After obtaining the loan, the borrower is likely to default unilaterally. Pope and Sydnor [15] also found it difficult to judge the risk of the personal information provided by the borrower unilaterally because the authenticity of this information cannot be verified. Freedman and Jin [16] uncovered that the borrower may commit fraudulent behavior by reporting false information, which exacerbates the information asymmetry between the two parties. Herzenstein *et al.* [17] also found that the borrowers' repayment ability and credit rating are the factors that have the greatest impact on personal credit risk. They concluded that economic strength is the determinant of judging the availability of borrowing. At the same time, Herzenstein *et al.* [18] depicted the borrowers' spending power can also directly affect the success rate of borrowing. These methods reveal the characteristics of the borrowers would be helpful for fraud detection.

Motivated by such an idea, we propose a deep learning technique to mine the fraud in a public lending dataset with 200,000 records. We analyze the customer credit rating, which can help us to identify customers' actual situations. Intuitively, the lower a customer has a credit rating, such as the E rating, the greater the likelihood of being a fraudulent user. Internet finance small loan companies set different thresholds on their customer credit rating data to build anti-fraud rules based on the true information of their customers. This paper aims to provide small financial credit companies a simple yet effective model to improve their risk control and the level of anti-fraud. Such companies often have a poor-risk control capacity with limited capacity for data engineering, modeling, and optimization. The main contribution of this paper is summarized as follows.

- First, we analyze the real-world Internet financial data for the missing data and sample imbalance. We propose to fill the missing with a random forest and deal with the sample imbalance with a synthetic minority oversampling technique.
- We train a deep neural network by the preprocessed data. We make comprehensible experiments for the setting of the network architecture and hyperparameters.
- Extensive experiments have been conducted to demonstrate the outperformance comparing with the commonly-used loan fraud detection models.

The rest of this paper is organized as follows. The second part is the methodology, and the third part is an empirical study from real-world data. The fifth part is the conclusion.

II. METHOD

A. INTERNET FINANCIAL FRAUD DETECTION

The reform of Internet technology has undoubtedly brought vitality to the development of the financial industry. With the emergence of various Internet finance models, the risks of Internet finance are also increasing. From the perspective of fraud, Internet financial fraud mainly has the following common types [19], [20], including:

- **Identity theft:** Criminals steal user's personal financial information in order to conduct fraudulent financial transaction activities or withdraw money from your account.
- **Investment fraud:** Selling investment or securities with false, misleading, or fraudulent information.
- **Mortgage and loan fraud:** The borrower uses false information to open a mortgage or loan, or the lender uses a high-pressure sales strategy to sell the mortgage or loan or predatory loan to users.
- **Large-scale marketing fraud:** Criminals usually use a lot of mail, telephone, or spam to steal users' personal financial information or request donations and fees from fraudulent organizations, usually involving fake checks, charities, sweepstakes, lotteries, and exclusive clubs or honor society invites.

B. DATA CLEANING AND FEATURE SELECTION

Data cleaning and feature selection are important prerequisites for building a featured anti-fraud model. If this part of the work is well processed, the model recognition efficiency will be greatly improved. We first delete the invalid variables or the variables with the same value that is unreasonable. Then, we figure out the missing values and mark up these holds to be filled. Next, we use a random forest [21] to experientially expand and derive the missing variables. Then, we employ an XGBoost model [22] to select

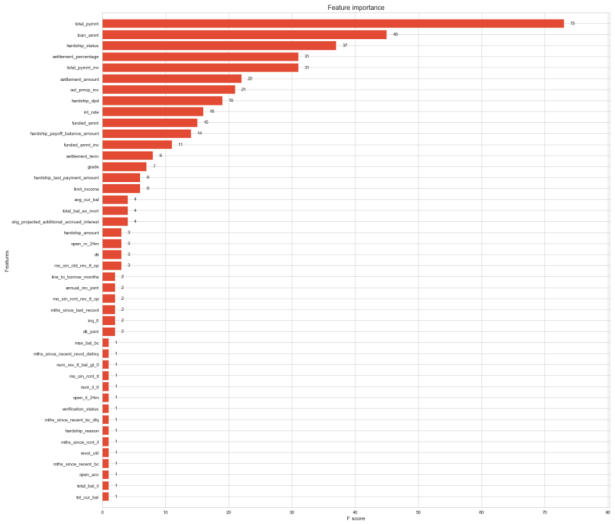

FIGURE 1. The feature importance calculated by XGBoost algorithm.

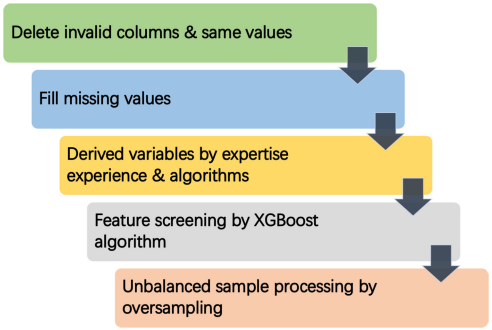

FIGURE 2. The flow chart for data preprocessing.

principle features. This results in 27 independent variable features and 1 target variable. The importance of the features is shown in Fig. 1 for example.

After selecting the features, the samples in each category are seriously imbalanced. In this regard, we introduce an oversampling synthetic minority oversampling technique (SMOTE) [23] for processing. Then, we normalize the processed data for modeling. The overall flowchart is shown in Fig. 2.

C. DEEP NEURAL NETWORK FOR INTERNET FRAUD DETECTION

After properly screen out the discriminated features, we formulate the Internet fraud detection task by a simple yet efficient deep neural network. The deep neural network (DNN) consists of an input layer, a few hidden layers, and an output layer. The first layer is the input layer, the last layer is the output layer, and the middle layers are all hidden layers [24]. Each layer contains multiple neurons. The number of neurons in the input layer depends on the input data. The number of neurons in other layers will be adjusted according to the actual situation. The number of hidden layers is customizable, often more than one layer. Layers are often fully connected,

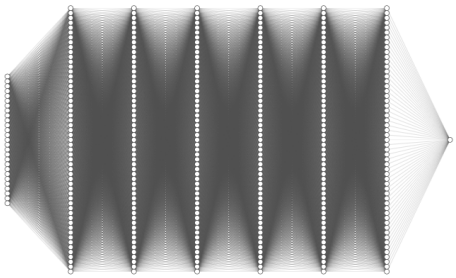

FIGURE 3. The architecture of the deep neural network.

that is, any neuron in each layer is connected to all neurons in the next layer. There is no connection between the neurons in the layer. Fig. 3 examples the deep neural network, which contains 3 hidden layers.

The deep neural networks achieve *self-learning* through forwarding propagation and backpropagation. Forward propagation is the process of feeding samples into the neural network, passing through the hidden layer, and finally getting the results from the output layer. The fitting degree of the model can be evaluated by the result of the loss function. Generally, the mean square error of the output layer result and the sample label is used as the loss function. The gradient descent method is commonly used in backpropagation to minimize the iterative optimization of the loss function. During this process, parameters such as weights and offset values are continuously updated, and the value of the loss function is constantly changing.

III. EXPERIMENTS

A. DATA DESCRIPTION

The experimental data comes from a large public lending dataset released by Lending Club. The data were collected from the fourth quarter of 2016 to the second quarter of 2017, with a total of 200,000 data records. Each record has 145 features, including 107 floating-point values and 38 character values. The initial dataset is 228.1 MB in size.

The samples in the dataset are separated into two part, 70% of samples were used for training, and the rest ones were used for testing. We select discriminate features by XGBoost [22]. Thirty selected features are used to train the Internet fraud detection model. All the feature data is normalized firstly. Then, these features are fed to a deep neural network, which contains an input layer, six hidden layers, and an output layer.

We list the most important variables and their interpretation in Table 1.

B. EMPIRICAL ANALYSIS

The distribution of loan amount is shown in Fig. 4. We see that the minimum loan amount is \$1,000 and the maximum is \$40,000. The loan amount is mainly concentrated around \$10,000, and the median is \$12,000. This company mainly

TABLE 1. Variable explanation for the dataset.

Variable name	Description
grade	Risk level
verification_status	Whether the source of income is verified
acc_open_past_24mths	Number of transactions in the past 24 months
inq_last_12m	Inquiries in the last 12 months
open_il_12m	Inquiries in the last 12 months
mths_since_rcnt_il	Number of accounts opened in the past 12 months
mo_sin_rcnt_tl	Number of installment accounts opened in the past 12 months
dti	Months since the opening of the account
all_util	Months since the opening of the account
il_util	Debt ratio
mo_sin_rcnt_rev_tl_op	Total debt (total credit balance)
inq_fi	Credit line / installment credit line
home_ownership	The number of months of the most recent revolving account (credit card)
mor_acc	Number of personal inquiries
tot_cur_bal	Nature of housing
total_rev_hi_lim	Number of mortgage accounts
acc_now_delinq	Total cash in all accounts

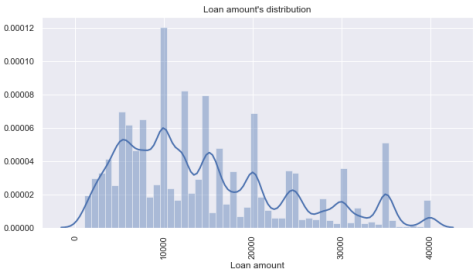

FIGURE 4. The distribution of loan amount.

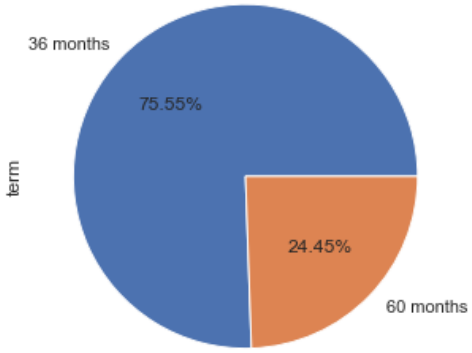

FIGURE 5. The distribution of loan term.

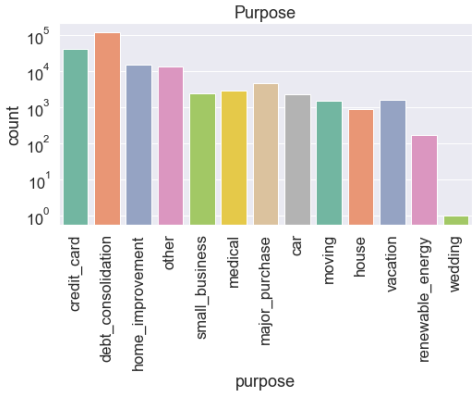

FIGURE 6. The distribution of loan usage.

focuses on small loans. Intuitively, the larger the loan amount there is, the greater the risk is.

The loan term distribution is shown in Fig. 5. In this dataset, there are two types of loan products, *e.g.* 36-month and 60-month. The proportion of loans with a 60-month loan is 24.45%, and the proportion of loans with a 36-month loan is 75.55%. As a common sense, the longer the loan term is, the higher the risk is. The greater the probability of default is, and the higher the risk of loan products with longer terms is. From a maturity perspective, assets with low risks are account for the majority.

The comparison of loan usage types is shown in Fig. 6. It can be seen that the most used loan is debt restructuring, which means one borrows new debt to repay the old debt. The credit card repayment is the second. One whose loan usage is for debt restructuring and credit card repayment has tight cash flow. Such customer is also unable to make loans before transferring to P2P platform loans. These customers have weak loan repayment ability. The possibility of default is higher. For other loans, the risk needs to be further analyzed through other characteristics of the customers.

The credit rating of the customers is shown in Fig. 7. The Lending Club divides customers' credit rating into 7 categories, *e.g.* A to G. The customers with a credit rating of A have the highest credit scores. The customers with a credit rating of G have the lowest ones. The customers with a high credit rating are less likely to default. At present, the customers with the most credit rating are Category B, followed by Category C and Category A. These three categories totally account for 81.12%. In addition, the customers with credit ratings of E, F, and G account for 5.42%. The credit department of Lending Club has stricter control over the credit status of applicants.

Finally, we demonstrate the relationship between loan usage and interest rate in Fig. 8.

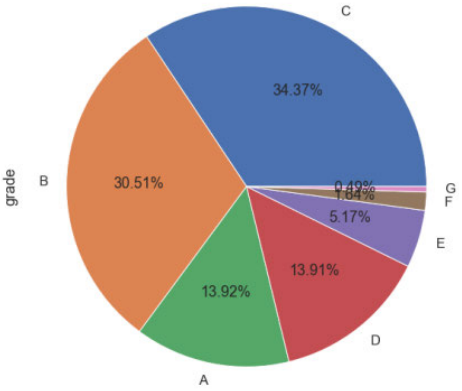

FIGURE 7. The proportion of customer credit rating.

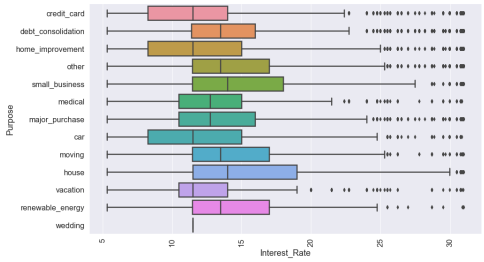

FIGURE 8. The relationship between the loan usage and interest rate.

TABLE 2. The model accuracy regarding to different combinations of batch size and epoches.

Batch size / Epoch	10	50	100	200	300
100	95.58%	96.07%	96.56%	97.18%	97.08%
500	95.21%	95.88%	96.49%	96.96%	96.82%
1000	95.02%	95.64%	96.37%	**97.61%**	96.67%
2000	94.29%	95.23%	96.11%	96.68%	96.91%
3000	94.06%	95.16%	95.95%	96.50%	96.76%
4000	93.72%	94.65%	95.72%	96.42%	96.72%

C. DESIGN DECISIONS

We employ the Tensorflow framework to implement the models mentioned in this paper. The number of nodes in the hidden layers is important. We use the following empirical formula to determine the nodes in the hidden layers.

$$N_h = \frac{N_s}{(\alpha \times (N_i + N_o))} \tag{1}$$

where N_i is the number of neurons in the input layers, N_o is the number of neurons in the output layer, N_s is the number of samples in the training set, α is often set from 2 to 10.

The batch size and epochs are important to calibrate the deep learning model. The following table shows the model accuracy corresponding to different combinations of the batch size and the epochs. We can see from Table 2, when batch size is 1000 and epoch is 200, the model obtains the highest accuracy. We set the batch size to 1000, and the epochs to 200 to train our model for performance evaluation.

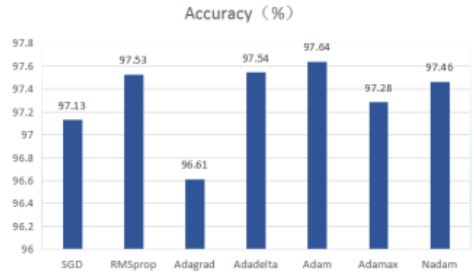

FIGURE 9. The accuracy of different commonly used optimizers.

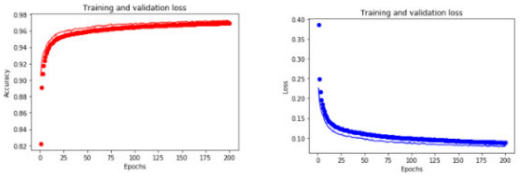

FIGURE 10. The accuracy and loss on the training and validation sets.

We compare various optimization algorithms to train the deep neural network. These optimization algorithms are commonly used in deep learning, *e.g.* SGD, RMSprop, AdaGrad, Adadelta, Adam, Adamax, and Nadam. The accuracy of different optimization algorithms corresponding to the DNN model is shown in Fig. 9.

We can see from Fig. 9, the performance of optimizers are similar, except the AdaGrad optimizer, whose accuracy rate is lower. The Adam optimizer is slightly better than the other six optimizers. Therefore, we adopt the Adam optimizer to train the DNN model.

D. EXPERIMENTAL RESULTS

We set the ReLU function and the sigmoid function as the activation function. The batch size is set to 1000, the number of epoch is set to 200, and the optimizer is set as the Adam function. We set the loss function of the model as binary cross-entropy.

As shown in Fig. 10, the accuracy change subtly after 175 iterations. The accuracy of the model is relatively high after 175 epochs. At this time, the loss is also relatively low. The final training result is the training accuracy of 0.9763 and the training loss of 0.0835. The testing accuracy is 0.9771 and the loss is 0.0789, which demonstrates the generalization ability of the proposed model is relatively good.

We evaluate the model by AUC and KS values, *e.g.* $AUC = 0.97$ and $KS = 0.94$ respectively, as shown in Fig. 11.

From Fig. 11, we see that the model has a generalization ability with high model stability.

E. PERFORMANCE EVALUATION AND DISCUSSIONS

We evaluate the performance of the deep neural network by comparing with four commonly-used models, *e.g.* logistic regression (LS), support vector machine (SVM), decision

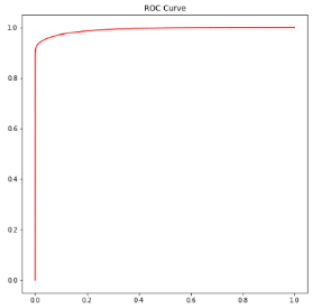

FIGURE 11. The ROC of the deep neural network.

TABLE 3. The comparison with commonly-used models regarding to AUC, KS, and ACC.

Model	AUC	KS	ACC
Decision Tree	0.9703	0.8987	0.9703
Support Vector Machine	0.9711	0.9149	0.9711
Logistic Regression	0.9719	0.8524	0.9720
Random Forest	0.9714	0.8762	0.9714
Deep Neural Network	0.9721	0.9442	0.9771

tree (DT), and random forest (RF). The evaluation results include AUC, KS, and ACC are listed in Table 3.

We can see from Table 3 that the deep neural network outperforms the comparisons, although the metrics are similar. The AUC of the decision tree is the lowest, the KS level is medium, and the ACC is the lowest. The results are slightly inferior to other models. This indicates the decision tree may degrade by the subtle noises inside the financial data. The AUC of logistic regression is second high, with the lowest KS and high ACC. This indicates that logistic regression has relatively good performance. However, its KS value is lower, which indicates that the optimal division degree is not as high as that of other models. The SVM algorithm has the lowest AUC, and the other two index values are at the intermediate level. The three evaluation index values of the random forest are at the intermediate level. The DNN model has the highest KS, the ACC is also the highest, and the AUC value is also the best. From Tables 3, the AUC, KS, and ACC of the DNN model are the best compared to the commonly-used models. This reveals the deep neural network can be applied for Internet fraud detection. The deep neural network is general, and can be extended to other applications, such as traffic flow forecasting [25]–[36], recommendation systems [37]–[39], medical image processing [40]–[44], intelligent computing [45]–[50].

IV. CONCLUSION

In this paper, we take the real customer information of the public loan data set of the lending club company as a sample. Then, we build a deep learning based Internet fraud detection model. We introduce the main parameters of the model and optimizes to find the optimal parameter combination of the model. Finally, the most popular logistic regression in the financial industry as well as other comparisons are used as a baseline to evaluate the performance of the proposed model. The results reveal the deep neural network achieves better performance, which is promising to be used in the financial industry for Internet fraud detection. In the future, we plan to cooperate with mature Internet financial technology companies and banks in China for blacklists and whitelists. The deep neural network combined with such blacklists and whitelists and the expertise anti-fraud rules is promising to increase fraud detection capability.

DECLARATION OF COMPETING INTEREST

The authors declare there are no conflicts of interest regarding the publication of this paper.

REFERENCES

[1] K. J. Leonard, "The development of a rule based expert system model for fraud alert in consumer credit," *Eur. J. Oper. Res.*, vol. 80, no. 2, pp. 350–356, Jan. 1995.
[2] D. Sánchez, M. A. Vila, L. Cerda, and J. M. Serrano, "Association rules applied to credit card fraud detection," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3630–3640, Mar. 2009.
[3] M. E. Edge and P. R. F. Sampaio, "The design of FFML: A rule-based policy modelling language for proactive fraud management in financial data streams," *Expert Syst. Appl.*, vol. 39, no. 11, pp. 9966–9985, Sep. 2012.
[4] S. Ghosh and D. L. Reilly, *Credit Card Fraud Detection With a Neural-Network*. Wailea, HI, USA: IEEE, 1994.
[5] A. I. Kokkinaki, "On atypical database transactions: Identification of probable frauds using machine learning for user profiling," in *Proc. IEEE Knowl. Data Eng. Exchange Workshop*, 1997, pp. 229–238.
[6] Y. Peng, G. Wang, G. Kou, and Y. Shi, "An empirical study of classification algorithm evaluation for financial risk prediction," *Appl. Soft Comput.*, vol. 11, no. 2, pp. 2906–2915, Mar. 2011.
[7] J. Z. Lei and A. A. Ghorbani, "Improved competitive learning neural networks for network intrusion and fraud detection," *Neurocomputing*, vol. 75, no. 1, pp. 135–145, Jan. 2012.
[8] Y. Sahin, S. Bulkan, and E. Duman, "A cost-sensitive decision tree approach for fraud detection," *Expert Syst. Appl.*, vol. 40, no. 15, pp. 5916–5923, Nov. 2013.
[9] N. Soltani Halvaiee and M. K. Akbari, "A novel model for credit card fraud detection using artificial immune systems," *Appl. Soft Comput.*, vol. 24, pp. 40–49, Nov. 2014.
[10] F. Louzada and A. Ara, "Bagging k-dependence probabilistic networks: An alternative powerful fraud detection tool," *Expert Syst. Appl.*, vol. 39, no. 14, pp. 11583–11592, Oct. 2012.
[11] M. Carminati, R. Caron, F. Maggi, I. Epifani, and S. Zanero, "BankSealer: A decision support system for online banking fraud analysis and investigation," *Comput. Secur.*, vol. 53, pp. 175–186, Sep. 2015.
[12] K. Fu, D. Cheng, Y. Tu, and L. Zhang, "Credit card fraud detection using convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process.*, Oct. 2016, pp. 483–490.
[13] B. Tu, D. He, Y. Shang, C. Zhou, and W. Li, "Deep feature representation for anti-fraud system," *J. Vis. Commun. Image Represent.*, vol. 59, pp. 253–256, Feb. 2019.
[14] M. E. Greiner and H. Wang, "Building consumer-to-consumer trust in E-Finance marketplaces: An empirical analysis," *Int. J. Electron. Commerce*, vol. 15, no. 2, pp. 105–136, Dec. 2010.
[15] D. G. Pope and J. R. Sydnor, "What's in a picture? Evidence of discrimination from prosper.com," *J. Hum. Resour.*, vol. 46, no. 1, pp. 53–92, 2011.
[16] S. Freedman and G. Z. Jin, "The information value of online social networks: Lessons from peer-to-peer lending," *Int. J. Ind. Org.*, vol. 51, pp. 185–222, Mar. 2017.
[17] M. Herzenstein, U. M. Dholakia, and R. L. Andrews, "Strategic herding Behavior in peer-to-peer loan auctions," *J. Interact. Marketing*, vol. 25, no. 1, pp. 27–36, Feb. 2011.
[18] M. Herzenstein, R. L. Andrews, and U. M. Dholakia, "The democratization of personal consumerloans? Determinants of successin online peer-to-peer lending communities," *J. Marketing Res.*, vol. 15, pp. 274–277, 2008.

[19] N. Mykhalchenko and J. Wiegratz, "Anti-fraud measures in Southern Africa," *Rev. Afr. Political Economy*, vol. 46, no. 161, pp. 496–514, Jul. 2019.
[20] B. Riley, "Anti-fraud technologies: A business essential in the card industry," *Card Technol. Today*, vol. 19, no. 10, pp. 10–11, Oct. 2007.
[21] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[22] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.
[23] T. Pan, J. Zhao, W. Wu, and J. Yang, "Learning imbalanced datasets based on SMOTE and Gaussian distribution," *Inf. Sci.*, vol. 512, pp. 1214–1233, Feb. 2020.
[24] E. Micheu-Tzanakou, "Artificial neural networks: An overview," *Netw.*, vol. 22, nos. 1–4, pp, pp. 208–230, 2011.
[25] T. Zhou, G. Han, X. Xu, Z. Lin, C. Han, Y. Huang, and J. Qin, "δ-agree AdaBoost stacked autoencoder for short-term traffic flow forecasting," *Neurocomputing*, vol. 247, pp. 31–38, Jul. 2017.
[26] T. Zhou, G. Han, X. Xu, C. Han, Y. Huang, and J. Qin, "A learning-based multimodel integrated framework for dynamic traffic flow forecasting," *Neural Process. Lett.*, vol. 49, no. 1, pp. 407–430, Feb. 2019.
[27] T. Zhou, D. Jiang, Z. Lin, G. Han, X. Xu, and J. Qin, "Hybrid dual Kalman filtering model for short-term traffic flow forecasting," *IET Intell. Transp. Syst.*, vol. 13, no. 6, pp. 1023–1032, Jun. 2019.
[28] L. Cai, Q. Chen, W. Cai, X. Xu, T. Zhou, and J. Qin, "SVRGSA: A hybrid learning based model for short-term traffic flow forecasting," *IET Intell. Transp. Syst.*, vol. 13, no. 9, pp. 1348–1355, Sep. 2019.
[29] W. Cai, D. Yu, Z. Wu, X. Du, and T. Zhou, "A hybrid ensemble learning framework for basketball outcomes prediction," *Phys. A, Stat. Mech. Appl.*, vol. 528, Aug. 2019, Art. no. 121461.
[30] S. Zhang, Y. Song, D. Jiang, T. Zhou, and J. Qin, "Noise-identified Kalman filter for short-term traffic flow forecasting," in *Proc. 15th Int. Conf. Mobile Ad-Hoc Sensor Netw. (MSN)*, Dec. 2019, pp. 1–5.
[31] W. Cai, J. Yang, Y. Yu, Y. Song, T. Zhou, and J. Qin, "PSO-ELM: A hybrid learning model for short-term traffic flow forecasting," *IEEE Access*, vol. 8, pp. 6505–6514, 2020.
[32] L. Cai, Z. Zhang, J. Yang, Y. Yu, T. Zhou, and J. Qin, "A noise-immune Kalman filter for short-term traffic flow forecasting," *Phys. A, Stat. Mech. Appl.*, vol. 536, pp. 1–9, Dec. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378437119314876
[33] H. Lu, Z. Ge, Y. Song, D. Jiang, T. Zhou, and J. Qin, "A temporal-aware LSTM enhanced by loss-switch mechanism for traffic flow forecasting," *Neurocomputing*, vol. 427, pp. 169–178, Feb. 2021.
[34] L. Cai, Y. Yu, S. Zhang, Y. Song, Z. Xiong, and T. Zhou, "A sample-rebalanced outlier-rejected k-nearest neighbor regression model for short-term traffic flow forecasting," *IEEE Access*, vol. 8, pp. 22686–22696, 2020.
[35] L. Cai, M. Lei, S. Zhang, Y. Yu, T. Zhou, and J. Qin, "A noise-immune lstm network for short-term traffic flow forecasting," *Chaos*, vol. 30, no. 3, pp. 1–10, 2020.
[36] H. Lu, D. Huang, Y. Song, D. Jiang, T. Zhou, and J. Qin, "ST-TrafficNet: A spatial-temporal deep learning network for traffic forecasting," *Electronics*, vol. 9, no. 9, p. 1474, Sep. 2020.
[37] D. Jiang, Z. Liu, L. Zheng, and J. Chen, "Factorization meets neural networks: A scalable and efficient recommender for solving the new user problem," *IEEE Access*, pp. 1–12, 2020.
[38] L. Zheng, N. Guo, W. Chen, J. Yu, and D. Jiang, "Sentiment-guided sequential recommendation," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2020, pp. 1957–1960.
[39] L. Zheng, N. Guo, J. Yu, and D. Jiang, "Memory reorganization: A symmetric memory network for reorganizing neighbors and topics to complete rating prediction," *IEEE Access*, vol. 8, pp. 81876–81886, 2020.
[40] T. Zhou, G. Han, B. N. Li, Z. Lin, E. J. Ciaccio, P. H. Green, and J. Qin, "Quantitative analysis of patients with celiac disease by video capsule endoscopy: A deep learning method," *Comput. Biol. Med.*, vol. 85, pp. 1–6, Jun. 2017.
[41] Y. Song, Z. Yu, T. Zhou, J. Y.-C. Teoh, B. Lei, K.-S. Choi, and J. Qin, "CNN in CT image segmentation: Beyond loss function for exploiting ground truth images," in *Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2020, pp. 1–4.
[42] D. Jiang, K. Wu, D. Chen, G. Tu, T. Zhou, A. Garg, and L. Gao, "A probability and integrated learning based classification algorithm for high-level human emotion recognition problems," *Measurement*, vol. 150, pp. 1–11, Jan. 2019.
[43] B. N. N. Li, X. Wang, R. Wang, T. Zhou, R. Gao, E. J. Ciaccio, and P. H. Green, "Celiac disease detection from videocapsule endoscopy images using strip principal component analysis," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, early access, Nov. 15, 2020, doi: 10.1109/TCBB.2019.2953701.
[44] X. Li, L. Bai, Z. Ge, Z. Lin, X. Yang, and T. Zhou, "Early diagnosis of neuropsychiatric systemic lupus erythematosus by deep learning enhanced magnetic resonance spectroscopy," *J. Med. Imag. Health Informat.*, vol. 11, no. 5, May 2021.
[45] J. Wang, Z. Xie, Y. Li, Y. Song, J. Yan, W. Bai, T. Zhou, and J. Qin, "Relationship between health status and physical fitness of college students from south China: An empirical study by data mining approach," *IEEE Access*, vol. 8, pp. 67466–67473, 2020.
[46] C. Li, S. Tang, H. K. Kwan, J. Yan, and T. Zhou, "Color correction based on CFA and enhancement based on retinex with dense pixels for underwater images," *IEEE Access*, vol. 8, pp. 155732–155741, 2020.
[47] C. Li, S. Tang, J. Yan, and T. Zhou, "Low-light image enhancement via pair of complementary gamma functions by fusion," *IEEE Access*, vol. 8, pp. 169887–169896, 2020.
[48] C. Li, S. Tang, J. Yan, and T. Zhou, "Low-light image enhancement based on quasi-symmetric correction functions by fusion," *Symmetry*, vol. 12, no. 9, p. 1561, Sep. 2020.
[49] G. Xiao, G. Tu, L. Zheng, T. Zhou, X. Li, S. H. Ahmed, and D. Jiang, "Multi-modality sentiment analysis in social Internet of Things based on hierarchical attentions and CSATTCN with MBM network," *IEEE Internet Things J.*, early access, Aug. 10, 2021, doi: 10.1109/JIOT.2020.3015381.
[50] D. Jiang, G. Tu, D. Jin, K. Wu, C. Liu, L. Zheng, and T. Zhou, "A hybrid intelligent model for acute hypotensive episode prediction with large-scale data," *Inf. Sci.*, vol. 546, pp. 787–802, Feb. 2021.

WEIWEI FANG is a Postdoctoral Research Fellow with the College of Science, Shantou University, China. Her research interests include financial analysis and accounting.

XIN LI is an Associate Professor with the Department of Computer Science, Shantou University. His research interests include intelligent transportation systems and machine learning.

PING ZHOU is currently pursuing the master's degree with the Department of Computer Science, College of Engineering, Shantou University, China. Her research interests include internet finance and deep learning.

JINGWEN YAN is a Full Professor with the College of Engineering, Shantou University, China. His research interests include machine learning, computer vision, and remote sensing.

DAZHI JIANG received the B.A. degree in computer science from the China University of Geosciences, Wuhan, in 2004, and the Ph.D. degree from the State Key Laboratory of Software Engineering, Wuhan University, China, in 2009. He was a Professor with the Department of Computer Science, Shantou University, China. Since 2009, he has been with the Department of Computer Science, Shantou University. His research interests include affective computing, deep learning, data mining, and applications of artificial intelligence.

TENG ZHOU is currently an Assistant Professor with the Department of Computer Science, Shantou University, and also a Research Associate with the Center of Smart Health, The Hong Kong Polytechnic University. His research interests include intelligent transportation systems and machine learning.