



# Machine learning-driven credit risk: a systemic review

Si Shi<sup>1</sup> · Rita Tse<sup>1,2</sup> · Wuman Luo<sup>1</sup> · Stefano D'Addona<sup>3</sup> · Giovanni Pau<sup>4,5</sup>

Received: 12 February 2022 / Accepted: 26 May 2022 / Published online: 16 July 2022  
© The Author(s) 2022

## Abstract

Credit risk assessment is at the core of modern economies. Traditionally, it is measured by statistical methods and manual auditing. Recent advances in financial artificial intelligence stemmed from a new wave of machine learning (ML)-driven credit risk models that gained tremendous attention from both industry and academia. In this paper, we systematically review a series of major research contributions (76 papers) over the past eight years using statistical, machine learning and deep learning techniques to address the problems of credit risk. Specifically, we propose a novel classification methodology for ML-driven credit risk algorithms and their performance ranking using public datasets. We further discuss the challenges including data imbalance, dataset inconsistency, model transparency, and inadequate utilization of deep learning models. The results of our review show that: 1) most deep learning models outperform classic machine learning and statistical algorithms in credit risk estimation, and 2) ensemble methods provide higher accuracy compared with single models. Finally, we present summary tables in terms of datasets and proposed models.

**Keywords** Credit risk · Machine learning · Deep learning · Statistical learning

## 1 Introduction

Machine learning advances heavily affected industry and academia in the past decades, ultimately transforming people's daily life. Artificial Intelligence (AI) has been applied to almost every human activity, including pattern recognition, image classification, business, agriculture, transportation, and finance. This paper focuses on machine learning applied to finance and credit risk estimation. Modern financial systems rely on credit and trust. Credit risk is a fundamental parameter that measures and predicts the default probabilities of a debtor. The correct estimation of credit risk is paramount for the entire system. Failing in the credit risk estimation can lead to systemic failures such as the sub-prime crisis of 2008. Consequently, lenders devote large amounts of resources to predict the credit-worthiness of consumers and companies to develop appropriate lending strategies that minimize their risks. Historically, credit risk approaches use statistical methods such as Linear Discriminant Analysis [1] and Logistic Regression [2]. These methods, however, do not easily handle large datasets.

Advances in computing power and availability of large credit datasets paved the way to AI-Driven credit risk estimation algorithms such as traditional machine learning and deep learning. Conventional machine learning

---

✉ Giovanni Pau  
giovanni.pau@unibo.it

Si Shi  
si.shi@ipm.edu.mo

Rita Tse  
ritatse@ipm.edu.mo

Wuman Luo  
luowuman@ipm.edu.mo

Stefano D'Addona  
daddona@uniroma3.it

<sup>1</sup> Faculty of Applied Sciences, Macao Polytechnic University, Macao SAR, China

<sup>2</sup> Engineering Research Centre of Applied Technology on Machine Translation and Artificial Intelligence of Ministry of Education, Macao Polytechnic University, Macao SAR, China

<sup>3</sup> Department of Political Science, University of Roma Tre, Rome, Italy

<sup>4</sup> Department of Computer Science and Engineering, University of Bologna, Bologna, Italy

<sup>5</sup> UCLA Samueli Computer Science, University of California, Los Angeles, USA

techniques, e.g., k-Nearest Neighbor [3], Random Forest [4] and Support Vector Machines [5], are more effective and flexible than statistical methods. In particular, the vital branch of machine learning-deep learning techniques [6] applied to large credit risk data lake outperform their predecessors both in accuracy and efficiency.

This paper presents a systemic review of credit risk estimation algorithms. It analyzes both the major statistical approaches and AI-based techniques with a critical spirit. The aim is to provide a comprehensive overview of the current leading credit risk estimation technology, providing justification and connections between past and present works. This work proposes a novel taxonomy combining finance with machine learning techniques. In addition, this work ranks their performance in terms of accuracy and costs. This paper also discusses the challenges and possible solutions in terms of four aspects: data imbalance, dataset inconsistency, model transparency, and inadequate utilization of deep learning methods.

The remainder of the paper is organized as follows: the survey methodology will be discussed in Sect. 2. Section 3 introduces the principles of statistical learning, machine learning and deep learning. Section 4 analyzes credit risk-related applications in detail. In Sect. 5, presented algorithms are discussed and ranked by their performance against public datasets. Finally, results and current challenges are summarized in Sect. 6; while Sect. 7 concludes this work.

## 2 Survey methodology

### 2.1 Methodology

We applied PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Fig. 1) reviewing methodology in our paper. First, we adopted five searching platforms for our investigation: Google Scholar, ACM, IEEEExplore, Springerlink, and ScienceDirect. We used the keywords “machine learning” or “deep learning” combined with “credit risk” while searching. We got 2400 articles in total. Then, we applied a filtering algorithm considering the trade-off between publication year and citations to proceed. After removing 1400 duplicate records, 800 ineligible records, and 76 incomplete articles, we obtained 124 screened records. Based on the relevance, we excluded 24 articles less related to the topic. After manually checking whether the paper has clear evaluation metrics, we further excluded another 24 papers. Finally, we kept 76 studies in terms of the relevancy to the research topic, precision of evaluation metrics, publication time, and number of citations as our source of reviewing.

Figure 1 depicts the PRISMA flow diagram

### 2.2 Inclusion and exclusion criteria

In this paper, we select three inclusion criteria: (1) the relevance of research topic, (2) the precision of evaluation metrics, (3) the publication year and citations. Moreover, the papers will be excluded if they are duplicated, incomplete, too early, low-related with the topic, having no clear metrics or comparatively low citations.

We show the whole workflow of the selection process in Fig. 2.

### 2.3 The datasets and approaches of the reviewed articles

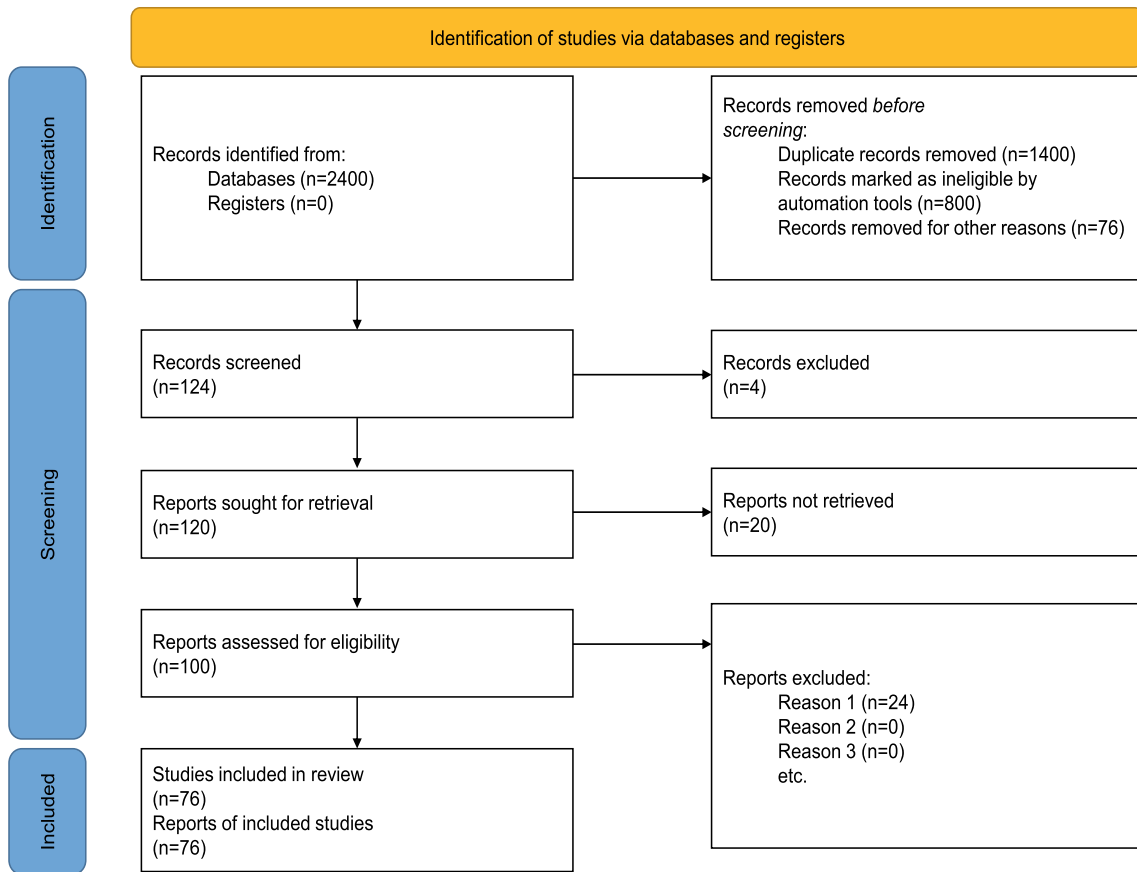
The mainly used datasets by the papers under review are German and Australian public credit data from the UCI Machine Learning repository [7–11]. In addition, there exist some researches that discover and mine their own data. For example, Chee Kian Leong (2015) uses data from a firm in Singapore [12]. Authors in [13–26] all employ their unique dataset. Those articles mainly emphasize the significance and the veracity of the original data.

We discuss the principles and application of the overall machine learning approaches. The traditional machine learning models for credit risk contain Support Vector Machines (SVMs) [5], k-Nearest Neighbor (k-NN) [3], Random Forests (RFs) [4], Decision Trees (DTs) [27–29], AdaBoost [30], Extreme Gradient Boost (XGBoost) [31], Stochastic Gradient Boosting (SGB) [32], Bagging [33], Extreme Learning Machine (ELM) [34] and GA (Genetic Algorithm) [35]. Neural network models generally belong to deep learning methods. Most of them include Convolutional Neural Networks (CNNs) [36], Deep Belief Neural Networks (DBNs) [37], Artificial Neural Networks (ANNs) [38], LSTM (Long Short-Term Memory) [39], Restricted Boltzmann Machines (RBMs) [40], Deep Multi-Layer Perceptron (DMLP) [41], and Recurrent Neural Networks (RNNs) [42].

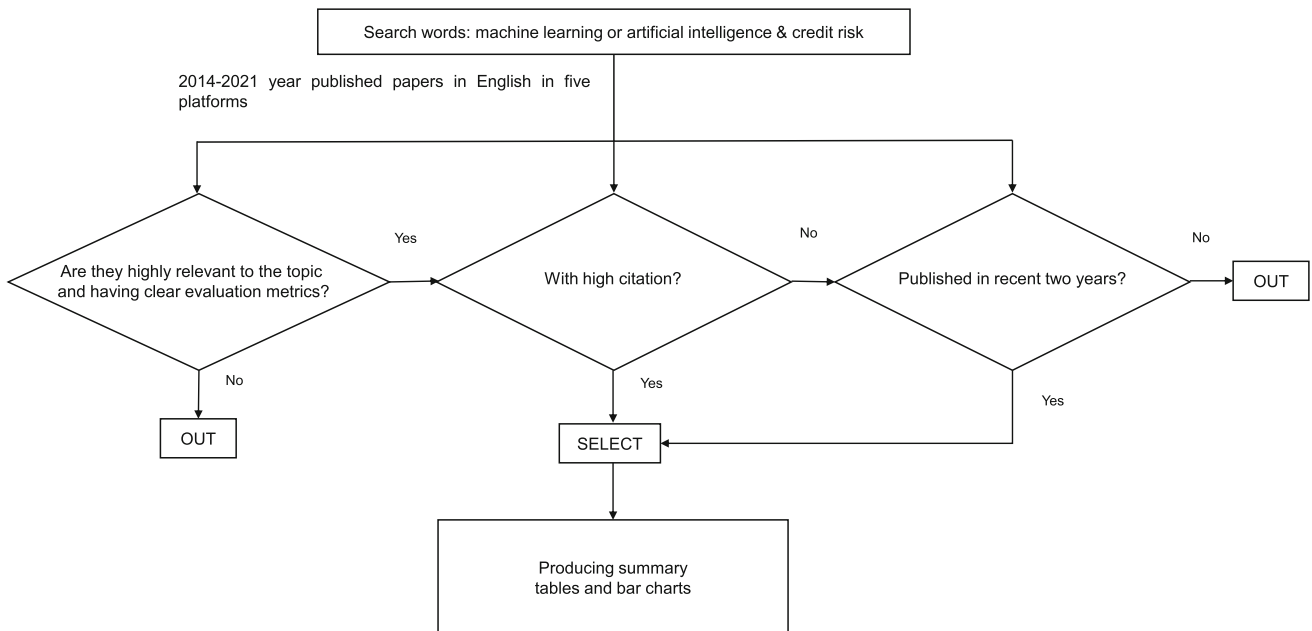
Summary tables and bar charts regarding all the methods of the reviewed papers are provided.

### 2.4 Taxonomy

The taxonomy is shown as Fig. 3. We can divide it into two parts: the first is regarding computing technology and the second is credit risk application domain. The two parts are further categorized into subsections. These two parts are connected and fused with each other. All the right-side subdomains include the left-side techniques, and all the techniques can be applied in the financial domains.



**Fig. 1** The PRISMA flow diagram. From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi:<https://doi.org/10.1136/bmj.n71>



**Fig. 2** The workflow of selecting papers

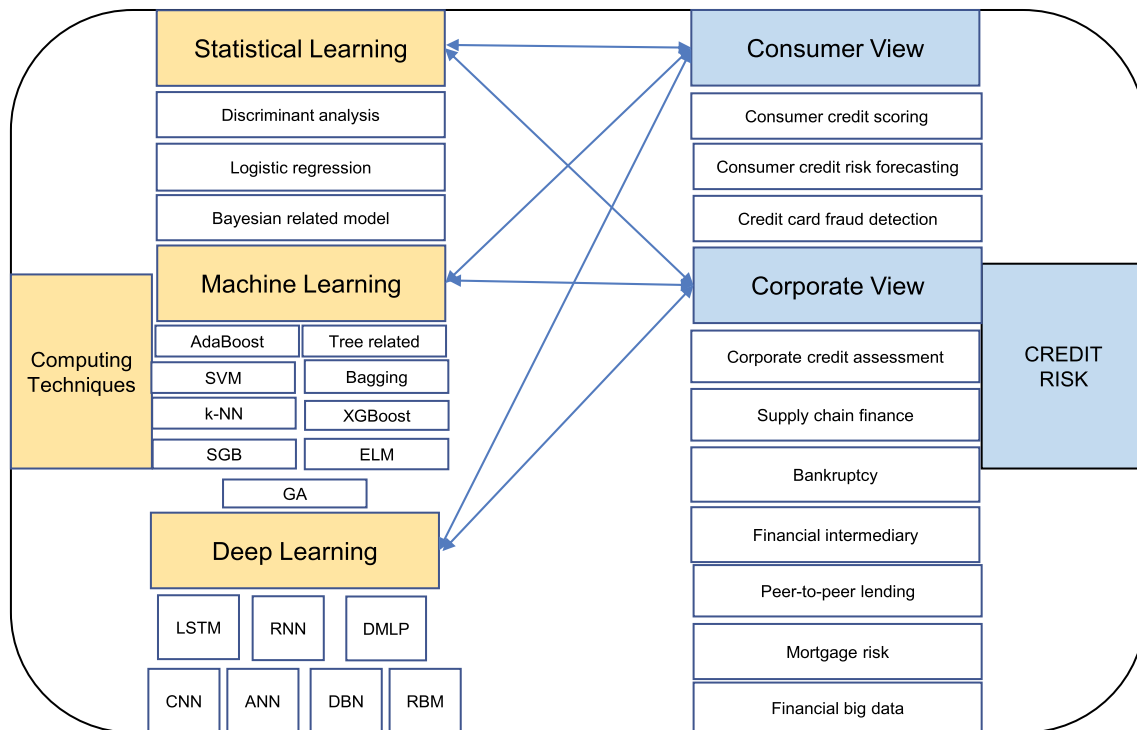


Fig. 3 The taxonomy of selecting paper

### 3 Computing approaches

This section briefly introduces three main computing techniques used for credit analysis, i.e., statistical learning, machine learning and deep learning, each of which has its own characteristics and similar principles. Statistical approaches are traditional ways to classify a customer's or enterprise's credit behavior. However, with the rapid development of artificial intelligence, machine learning and deep learning gradually took the place of statistical analysis.

#### 3.1 Statistical learning approaches

We divide the statistical approaches into three subsections—discriminant analysis, logistic regression and Bayesian related model.

LDA (Linear Discriminant Analysis) is a classic technique for predicting groups of samples [1]. It aims at generating characteristics that can separate binary variables.

Logistic regression is a classification algorithm which uses the logistic sigmoid function to squash the output of the linear function into the interval (0, 1) and interpret that value as a probability [6].

Naïve Bayes methods are statistical learning algorithms that apply Bayes' theorem with the “naïve” assumption of conditional independence between every pair of features if

the class variable is given [43]. A Bayesian network is a probabilistic model based on graphs. It measures the conditional dependence structure of a series of random variables that comply with the Bayes theorem [44].

#### 3.2 Machine learning methods

We review a series of conventional machine learning algorithms that can be applied well in credit risk area.

k-NN [3] belongs to classification methods that appoint the class of the majority of the  $k$  nearest neighbors of an input variable  $x$  to it in a dataset [3].

Tree-related methods show their effects in credit risk domain. Typical examples include DTs [27–29], Random Forests (RFs) [4], Classification and Regression Trees (CART) [45], C4.5 [46], and Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples (DECORATE) [47].

Support Vector Machine [5] implements a hyperplane (a decision boundary) which can separate classes in a high dimensional feature space. It outputs a class identity according to whether  $w^T + b$  is positive or not [6]. Here,  $w$  stands for the margin between the negative and positive hyperplane while  $b$  means the bias.

Boosting is an ensemble method that combines the individual models to gain higher capacity [6]. Adaptive Boosting (AdaBoost) belongs to the most popular boosting algorithms as the weights are re-assigned to each instance,

with higher weights assigned to incorrectly classified instances [48]. SGB (Stochastic Gradient Boosting) [32] can add incorporate randomness as an integral part when created from Gradient Boosting algorithm. This family of algorithms includes Extreme Gradient Boost (XGBoost) [31] similar to Gradient boosting. However, it includes the decision trees built in parallel rather than in a series manners.

Bagging is an ensemble method which contains the same kind of model, training algorithm, and objective function in recycling [6]. It is also known as bootstrap aggregation [33].

Extreme learning machine [34] was developed by Guang-Bin Huang in 2006. It targets at building single-hidden layer feedforward neural networks (SLFNs) which randomly chooses hidden nodes and outputs the weights of SLFNs logically [34].

Genetic algorithm (GA) [35] is a heuristic search algorithm to solve searching and optimization problems. It first generates an initial population, then obtains a fitness score for all individuals in it. Individuals are selected for the reproduction of offspring [49].

### 3.3 Deep learning methods

Deep learning has deeper layers and more units within a layer compared with traditional machine learning. It can represent functions of increasing complexity [6]. In this section, we review some crucial deep learning methods used in credit risk.

Artificial Neural Networks [38] were inspired by a biological neural network system. It has three layers generally: an input, hidden and output layers. Given a feature vector  $x$ , the ANN outputs  $\hat{y}$  through the following formula [50]:

$$\hat{y} = a_2(a_1(\alpha^{(1)}x + \alpha_0^{(1)})\alpha^{(2)} + \alpha_0^{(2)}) \quad (1)$$

where  $\alpha_0^{(1)}, \alpha^{(1)}, \alpha_0^{(2)}, \alpha^{(2)}$  are weights and  $a_1, a_2$  are activation functions.

Recurrent neural networks (RNNs) are a family of neural networks for processing sequential data [6]. They can better handle sequential information rather than the spatial data which Convolutional Neural Networks (CNNs) can effectively process. RNNs introduce state variables to store past information as well as the current inputs, both determining the current outputs [51].

LSTM [39] was first developed to produce paths in which the gradient flows for long durations [6]. It is the variant of Recurrent Neural Networks (RNNs). Compared with traditional RNNs, it can solve gradient disappearance and explosion in the long-term sequence process.

DMLP is a Multi-Layer Perceptron with multiple hidden layers. It is a directed neural network. In order to update the weights, the loss function for DMLP uses Softmax and Cross-Entropy. [50].

LeCun et al. first introduced CNNs [36] which were widely applied in image processing, voice recognition, automatic QA systems, and many other computing fields. CNNs consist of an input layer, convolutional layers, pooling layers and fully connected layers.

The convolution function is following [6]:

$$s(t) = \int_{-\infty}^{\infty} x(a)w(t-a) da \quad (2)$$

where  $w(a)$  is a weighing function where  $a$  is the age of a measurement.

Hinton et al. introduced DBNs [37] which are a class of deep neural networks. A typical DBN consists of several hidden layers of Restricted Boltzmann Machines. An output of a lower level RBM can be regarded as input of the higher level RBM [50].

RBM are some of the most common building blocks of deep probabilistic models. They are undirected probabilistic graphical models containing a layer of observable variables and a single layer of latent variables [6].

It has the similar energy function like Boltzmann Machine. The function is as follows [50]:

$$E(v, h) = - \sum_{i=1}^n a_i v_i - \sum_{j=1}^m b_j h_j - \sum_{i=1}^n \sum_{j=1}^m \alpha_{ij} v_i h_j \quad (3)$$

where  $a_i, b_j$  are biases for binary variables  $v_i, h_j$ , and  $\alpha_{ij}$  are weights between  $j$  and  $i$ .

## 4 Credit risk application with computing algorithms

In the past decades, a lot of scholars have employed various computing algorithms and models to solve credit risk prediction and assessment. Binary classification problem is the most fundamental and essential computing technique in credit risk scenarios. In this section, we divide the related studies into two groups from the perspective of finance: consumer and corporate.

### 4.1 Consumer credit risk

Consumer credit scoring is one of the main parts of credit risk management. It is a kind of system which determines the creditworthiness of a customer based on his/her past credit situation. In [52], the Bayesian network method is improved to find out whether there is a change in credit risk profiles. Numerous approaches have been implemented in

this domain. Typical examples include Extreme learning machines (ELM) [7], Ensemble of classifiers [8], Bayesian networks [12], Deep Genetic Cascade Ensembles [11], a hybrid model with convolutional neural networks and Relief algorithm [22], Genetic Programming [53], feature selection [54], RNN [55], ensemble of supervised learning and statistical learning [56], Radial Basis Function [57], TreeSHAP method for Stochastic Gradient Boosting [58], a real-time binary classification model [59], CNN [60], MLP [61], etc. The authors in [62] compared the traditional and machine learning models in the credit score evaluation area.

Predicting a consumer's future credit condition is also valuable for credit risk and quantitative analysis. The authors in [19] conduct a comparison between deep learning techniques and other machine learning methods. It proves that XGBoost overperforms traditional machine learning techniques like Logistic regression, SVM and Random Forest. It turns out that a hybrid model is capable of predicting credit risk. In [63], a unique model named TRUST (Trainable Undersampling with SELF Training) was proved to be decisive.

CCF (Credit card fraud) is a specific crime in the banking system and becoming a substantially growing problem worldwide [64]. Detection of it helps to control the credit risk in banking security issue. A novel framework called DEAL (Deep Ensemble Algorithm) is employed [64]. Recurrent Neural Network (RNN) [65], Boosted Decision Tree [66–68], a deep learning structure with an advanced feature engineering [69] display a satisfactory performance. The authors in [70] conduct a comparison among Deep Learning, Logistic Regression and Gradient Boosted Tree. In [71–73], the authors implemented LR, SVM, k-NN, NB, RF, DT, MLP methods and found that they were all robust while tree-related models have the best performance. By using an auto-encoder, the authors in [70] create features with domain expertise. It is proved to be an improvement in predictive power. In [74], Visual Analytics were used to help reduce the incidence of false positives.

## 4.2 Corporate credit risk

Credit risk in corporate aspect also demands the necessity of machine learning and deep learning.

Deep learning plays a significant role in corporate credit rating and assessment. Two-layer additive risk model [13], Artificial Neural Network [15], LSTM and AdaBoost [9], denoising-based neural network [21], deep belief network [14, 75], probabilistic neural network (PNN) [76], Genetic algorithm with neural network [77], CNN [78] all show their great competency in estimation and assessing.

Online supply chain financial risk can be controlled by proper estimation and assessment. The authors in [24] construct a deep belief network based on Restricted Boltzmann Machine and classifier SOFTMAX. The dataset came from annual financial reports of Chinese notable companies. The model shows an accuracy which is far beyond SVM and Logistic Regression. In [79, 80], SVM and XGBoost were more accurate than LR and NB in supply chain fraud detection.

Because of the remaining effect of Global Financial Crisis in 2008, a large number of corporations are under the threats of bankruptcy. Neural networks can help those in danger detect the early signals of collapsing. A series of machine learning methods are enforced to predict bankruptcy [16, 81]. Bagging, boosting and random forest have the best performance. In [10], random forest trees are proven to outperform most of the other machine learning models.

In [26], statistical methods—probit models and CART (Classification And Regression Trees), machine learning methods—Neural Networks and k-NN are applied and compared to make a prediction in financial intermediary domain.

International finance, which has an important branch peer-to-peer lending, once flourished in the past decades. Normally, it has greater credit risk than common financial industry. Neural Networks [17, 82, 83], Attention Mechanism LSTM [20], word embedding models [84], Ensemble Learning Method [85, 86], Restricted Boltzmann Machine (RBM) [87] all exert their impact on predicting the risk of P2P industry.

Mortgage credit and prepayment risk are vital issues for measuring a borrower's behavior in real estate financial industry. In [88], the authors find a highly nonlinear relationship between a borrower's behavior and risk factors with deep neural networks. Deep learning is proved to be effective in measuring mortgage risks.

Big data technology triggered the massive transformation of finance. According to Denis Ostapchenya, a financial expert, big data in banking can be deployed to assess risks in the procedure of trading stocks or checking the creditworthiness of a loan applicant. Big Data analysis also accelerates and ensures the processes which require compliance verification, auditing, and reporting [89]. In the credit risk domain, the combination of machine learning, big data and specific financial techniques has achieved satisfactory results. BP neural networks, genetic algorithm [90], logistic regression with XGBoost and AdaBoost [91, 92], Synthetic Minority Oversampling Technique algorithm [93], integrated and mixed models [94, 95] all play a vital role in predicting and classifying credit risk assessment.

## 5 Performance ranking of machine learning techniques

### 5.1 Data imbalance

Generally, data imbalance often occurs in the credit risk classification due to the huge differences of the number of good borrowers and bad borrowers. SMOTE [93] is one of the most widely used approach to address this problem. In addition, over-sampling and under-sampling techniques are also employed. Nevertheless, data imbalance has been severely underestimated in many credit risk researches.

### 5.2 Evaluation metrics

In this review, we select ACC (accuracy) and AUC as main metrics for performance evaluation. The metric accuracy (ACC) is calculated through correctly classified values divided by the total number of samples while the metric AUC is the area under ROC curve which is also a measurement of precision of classification.

ACC is calculated as follows:

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \tag{4}$$

where TP denotes true positive, TN stands for true negative, FP means false positive, FN denotes false negative.

AUC [96] can be expressed as the following formula:

$$AUC = \frac{1 + \frac{TP}{TP+FN} - \frac{FP}{FP+TN}}{2} \tag{5}$$

### 5.3 Ranking of techniques

There hasn't been consensus on the specific ranking of each machine learning technique. In this section, we propose our own thoughts that is based on a thorough and objective investigation. Because the open-source databases of German and Australian credit risk have uniform judging criteria, we select the common techniques appearing in the related literature to compare their performances. We use the mean of each metric of the methods. The bar charts are shown in Fig. 4.

The graph shows that machine learning methods have a higher accuracy universally than statistical methods. Bagging has the highest AUC and Random Forest (RF) has the highest ACC. Logistic Regression is the most powerful tool among the statistical methods in the credit risk classification. Naïve Bayes (NB), k-Nearest Neighbor (k-NN) and Classification and Regression Trees (CART) have comparatively low rankings regarding German credit dataset. The detailed ranking results are shown in Table 1.

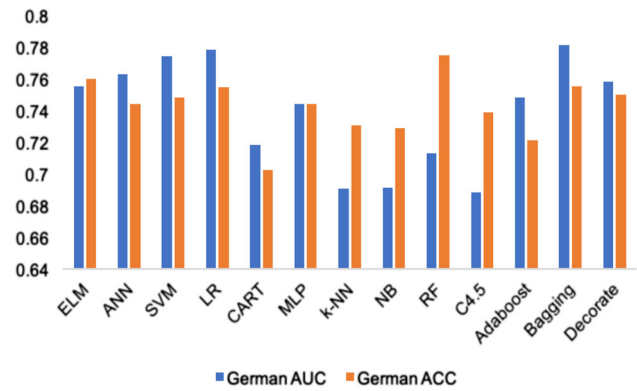


Fig. 4 The accuracy from German credit data

Table 1 Rank from German credit data

Methods	Rank according to AUC	Rank according to ACC
Bagging	1	3
LR	2	4
SVM	3	6
ANN	4	7
Decorate	5	5
ELM	6	2
AdaBoost	7	12
MLP	8	8
CART	9	13
RF	10	1
NB	11	11
k-NN	12	10
C4.5	13	9

All of the above methods are abbreviations for the notions introduced in the former sections

Similarly, we sort and calculate the mean ACC and AUC appearing in the Australian Credit Risk dataset. The result is shown in Fig. 5. It turns out that the accuracy in

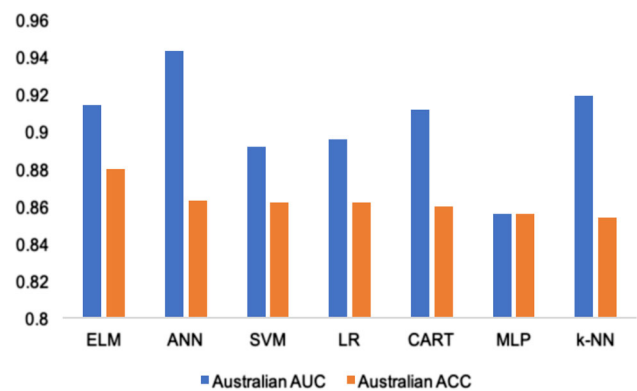


Fig. 5 The accuracy from Australian credit data

the Australian dataset exceeds the one in the German dataset because the imbalanced ratio of German dataset is comparatively higher. The best AUC is contributed by ANN. The best ACC belongs to ELM method.

From Fig. 5, we can conclude that deep learning methods are more potent than traditional machine learning and statistical methods from the above graph. The specific ranking is shown in Table 2.

In short, deep learning techniques have better performance regarding public credit risk data sets compared with machine learning and statistical learning methods based on ACC and AUC values.

## 6 Discussions

### 6.1 Existing survey papers

In this section, we review several typical surveys published recently. In [97], the majority of machine learning methods and data imbalance are discussed, but the discussion only focuses on the card defraud domain and the authors didn't consider the synergetic effects of models. Xolani Dastile, Turgay Celik et al. [50] had a thorough investigation of systematic machine learning and its application in credit risk. Nevertheless, the role of deep learning models in credit risk hasn't been fully expressed. In [98], principles of machine learning methods are not clearly displayed. In [99], abundant bibliography is shown. However, the structure of the paper is not balanced. Siddharth Bhatore et al. [100] displayed an intact review of machine learning in credit risk and showed clear graphs, but they ignored the limitation of datasets in some sense. In [101], similar problems with [98] occurred.

In our work, we give a comprehensive analysis and provide detailed comparison among methods, hoping to improve existing results.

**Table 2** Rank from Australian credit data

Methods	Rank according to AUC	Rank according to ACC
ANN	1	2
k-NN	2	7
ELM	3	1
CART	4	5
LR	5	4
SVM	6	3
MLP	7	6

All of the above methods are abbreviations for the notions introduced in the former sections

### 6.2 The summary tables

We summarize our survey in the following four tables. A whole summary table is shown in S1 Table.

Table 3 shows that LR and Bayesian models are the mostly used ones among the statistical learning techniques.

As shown in Table 4, we find that AdaBoost, SVM, Tree-related, k-NN and Bagging are the primarily implemented models among the machine learning techniques while SGB (Stochastic Gradient Boosting) and ELM (Extreme Learning Machine) have a relatively low citation.

Table 5 shows that ANN and MLP are the widely used deep learning models. Moreover, nearly all of the listed deep learning methods have a balanced citation distribution.

We list several important works containing unique datasets as Table 6. Almost all of them deploy their own computing models that improve the original algorithms. The results show that the models are effective.

### 6.3 Challenges

We summarize four major challenges in the research of machine learning-driven credit risk. First, data imbalance in credit risk is quite severe. Although several approaches such as over-sampling and under-sampling (usually chosen to under-sample the majority) have been proposed to solve this problem, the results are still unsatisfactory in terms of both effectiveness and efficiency. Second, the shortage of benchmark datasets is serious. Most existing works use private datasets, thus the results of performance comparison cannot be fair enough. Third, most machine learning models are black boxes since they are generally not transparent. Information transparency should be noticed.

**Table 3** Papers containing statistical learning models

Source	LDA	LR	Bayesian
[9]	+	+	+
[12]		+	+
[23]		+	+
[102]		+	
[92]		+	
[72]		+	+
[79]		+	+
[103]		+	+
[78]	+		+
[95]		+	+
[73]		+	+
[80]		+	+

LDA stands for linear discriminant analysis, LR stands for logistic regression



**Table 4** Papers containing machine learning models

Source	AdaBoost	SVM	Tree	kNN	Bagging	XGBoost	SGB	ELM	GA
[13]	+	+	+						
[81]	+		+		+		+		
[7]		+	+	+				+	
[8]	+		+		+				
[9]		+	+	+					
[10]	+		+		+				
[16]		+	+		+				
[18]		+	+	+					
[19]		+	+			+			
[11]		+							
[26]			+	+					
[53]									+
[76]			+						+
[90]									+
[93]	+	+	+						
[92]	+	+	+						
[72]		+	+	+					
[78]	+	+	+	+		+			

All of the above methods are abbreviations for the notions introduced in the former sections

**Table 5** Papers containing deep learning models

Source	CNN	MLP	ANN	LSTM	RBM	DBN	RNN
[14]		+				+	
[7]			+				
[9]			+	+			
[17]			+				
[12]			+				
[20]				+			
[21]			+				
[22]	+						
[75]			+			+	
[23]		+					
[24]					+		
[25]	+	+		+		+	
[26]			+				
[55]							+
[65]							+
[102]		+					
[87]					+		
[93]			+				
[71]		+					
[78]	+	+					
[60]	+						
[61]		+					

All of the above methods are abbreviations for the notions introduced in the former sections

Fourth, the application of deep learning models is still limited in credit risk.

These four challenges are what we are supposed to overcome in future work. We hope more and more deep advanced models will emerge in this area.

## 7 Conclusions

In conclusion, we have witnessed an overall application of machine learning as well as deep learning methods in credit risk area. We build a taxonomy which links computing algorithms and finance. We also briefly introduce the principles of statistical and machine learning approaches. As for public datasets, we rank them according to their accuracy. In addition, we list some of the accuracy for the private and unique datasets. A checklist is provided in S2 Table.

The results show that deep learning methods are more powerful than the traditional machine learning and statistical approaches although they haven't been fully employed. Also, the conclusion that ensembles of several methods outperform a single one has been proved in some of the related researches [9, 11, 75, 81, 103, 104].

In the future, we are supposed to find proper solutions to the challenges mentioned above. First, we should find new ways to tackle the problem of imbalanced data. Second, we will find a comprehensive judging criterion to make up for the default of specific methods and the inconsistency of

**Table 6** A summary table with unique datasets

Source	Datasets	Model	Accuracy
[13]	Balanced FICO (Fair Isaac Corporation)	Two-Layer Additive Risk Model	0.7404
[14]	Credit Default Swap	DBN	1
[15]	76 small businesses from a bank in Italy	Feedforward networks	0.87
[16]	NYU's Salomon Center database	Boosting	0.8631
[17]	A leading European P2P Platform Bondora	Neural networks	0.7438
[12]	An original credit scoring dataset of a firm providing credit and loans in Singapore	Neural networks	0.84
[18]	Lending club and kaggle, German credit	Balasso based RF	0.8975
[19]	A dataset belonging to National Bank of Canada	XGBoost	0.7822(Auc)
[20]	A real-world P2P Chinese data platform	AM LSTM	0.669(Auc)
[21]	Lending club	A denoising-autoencoder-based neural network model	0.875
[22]	A real-world dataset from a Chinese company	Relief-CNN	0.6989 (Auc)
[23]	The credit data from CBRC, German dataset, Darden dataset	FV-SMO	0.849 (Auc)
[24]	The database from Tai'an	DBN	0.9604
[25]	Bloomberg and Compustat	CNN2d	0.9013
[26]	Mortgage loan data from a commercial bank	CART	0.917

All of the above methods are abbreviations for the notions introduced in the former sections

datasets. Third, we should seek improvements in machine learning methods in tackling data transparency. Fourth, we should try our new and improved deep learning models in credit risk classification problem.

Moreover, in recent years, some authors proposed a series of representative nature-inspired metaheuristic algorithms such as (monarch butterfly optimization) MBO [105], (earthworm optimization algorithm) EOA [106], (elephant herding optimization) EHO [107], (moth search algorithm) MS [108], (Slime mould algorithm) SMA [109], (hunger games search) HGS [110], (colony predation algorithm) CPA [111] and (Harris hawks optimization) HHO [112]. They can also be applied in credit risk prediction. Besides, (Runge Kutta optimizer) RUN [113] is an algorithm that excludes the general characteristics of metaphor among other metaheuristic algorithms. Generally, those novel intelligent computational algorithms haven't been sufficiently applied in finance due to the complexity and instability of risk related problems. However, they may have promising results when the analysis tools become more mature.

Last but not least, big data technology and its application in credit risk is a newly booming area. We will explore them and utilize the vast amounts and efficiency of big data tools like MapReduce and Hadoop platform to get better results.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00521-022-07472-2>.

**Acknowledgements** This work was supported in part by the Macao Polytechnic University – Edge Sensing and Computing: Enabling Human-centric (Sustainable) Smart Cities (RP/ESCA-01/2020) and by the Emilia Romagna Region within the European S3 program with the Project LIBER. We want to thank the Macao government and the Emilia Romagna regional government for supporting this work.

**Author Contributions** Author Si Shi's contribution is to write and edit the whole paper. Author RT contribution is to provide the funding and resources as well as conceptualization of the paper. Author WL took part in the revising and conceptualization of the paper. Author SD'A helped with the formation of initial financial framework and revising. Author GP guided the whole research process and revised the paper.

**Funding** Open access funding provided by Alma Mater Studiorum - Università di Bologna within the CRUI-CARE Agreement. This work was funded in part by the Macao Polytechnic University – Edge Sensing and Computing: Enabling Human-centric (Sustainable) Smart Cities (RP/ESCA-01/2020).

**Availability of data and materials** Not applicable.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethics approval** The authors are consistent with the ethical requirements.

**Consent to participate** The authors all consent to participate in the paper editing.

**Consent for publication** The authors all consent to the publication of the paper.

**Code availability** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Moo-Young M (2019) Comprehensive biotechnology. Elsevier, Amsterdam
- Cox DR (1958) The regression analysis of binary sequences. *J R Stat Soc Ser B* 20(2):215–232
- Cover T, Hart P (1967) Nearest neighbor pattern classification. *IEEE Trans Inf Theory* 13(1):21–27
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297
- Goodfellow I, Bengio Y, Courville A (2016) Deep Learn. MIT press, Cambridge
- Bequé A, Lessmann S (2017) Extreme learning machines for credit scoring: An empirical evaluation. *Expert Syst Appl* 86:42–53
- Abellán J, Castellano JG (2017) A comparative study on base classifiers in ensemble methods for credit scoring. *Expert Syst Appl* 73:1–10
- Shen F, Zhao X, Kou G et al (2021) A new deep learning ensemble credit risk evaluation model with an improved synthetic minority oversampling technique. *Appl Soft Comput* 98(106):852
- Ghatasheh N (2014) Business analytics using random forest trees for credit risk prediction: a comparison study. *Int J Adv Sci Technol* 72(2014):19–30
- Plawiak P, Abdar M, Acharya UR (2019) Application of new deep genetic cascade ensemble of svm classifiers to predict the australian credit scoring. *Appl Soft Comput* 84(105):740
- Leong CK (2016) Credit risk scoring with bayesian network models. *Comput Econ* 47(3):423–446
- Chen C, Lin K, Rudin C, et al (2018) An interpretable model with globally consistent explanations for credit risk. arXiv preprint [arXiv:1811.12615](https://arxiv.org/abs/1811.12615)
- Luo C, Wu D, Wu D (2017) A deep learning approach for credit scoring using credit default swaps. *Eng Appl Artif Intell* 65:465–470
- Angelini E, Di Tollo G, Roli A (2008) A neural network approach for credit risk evaluation. *Quarte Rev Econ Finan* 48(4):733–755
- Barboza F, Kimura H, Altman E (2017) Machine learning models and bankruptcy prediction. *Expert Syst Appl* 83:405–417
- Byanjankar A, Heikkilä M, Mezei J (2015) Predicting credit risk in peer-to-peer lending: A neural network approach. In: 2015 IEEE symposium series on computational intelligence, IEEE, pp 719–725
- Arora N, Kaur PD (2020) A bolasso based consistent feature selection enabled random forest classification algorithm: an application to credit risk assessment. *Appl Soft Comput* 86(105):936
- Marceau L, Qiu L, Vandewiele N, et al (2019) A comparison of deep learning performances with other machine learning algorithms on credit scoring unbalanced data. arXiv preprint [arXiv:1907.12363](https://arxiv.org/abs/1907.12363)
- Wang C, Han D, Liu Q et al (2018) A deep learning approach for credit scoring of peer-to-peer lending using attention mechanism lstm. *IEEE Access* 7:2161–2168
- Fan Q, Yang J (2018) A denoising autoencoder approach for credit risk analysis. In: Proceedings of the 2018 international conference on computing and artificial intelligence, pp 62–65
- Zhu B, Yang W, Wang H, et al (2018) A hybrid deep learning model for consumer credit scoring. In: 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD), IEEE, pp 205–208
- Zhang Q, Wang J, Lu A et al (2018) An improved smo algorithm for financial credit risk assessment-evidence from china's banking. *Neurocomputing* 272:314–325
- Xu RZ, He MK (2020) Application of deep learning neural network in online supply chain financial credit risk assessment. In: 2020 international conference on computer information and big data applications (CIBDA), IEEE, pp 224–232
- Golbayani P, Wang D, Florescu I (2020) Application of deep neural networks to assess corporate credit rating. arXiv preprint [arXiv:2003.02334](https://arxiv.org/abs/2003.02334)
- Galindo J, Tamayo P (2000) Credit risk assessment using statistical and machine learning: basic methodology and risk modeling applications. *Comput Econ* 15(1):107–143
- Quinlan JR (1993) C4. 5: Programming for machine learning. *Morgan Kauffmann* 38(48):49
- Breimann L, Friedman JH, Olshen RA et al (1984) *Classif Regres Trees*. Wadsworth, Pacific Grove
- Quinlan JR (1986) Induction of decision trees. *Mach Learn* 1(1):81–106
- Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 55(1):119–139
- Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp 785–794
- Friedman JH (2002) Stochastic gradient boosting. *Comput Stat Data Anal* 38(4):367–378
- Breiman L (1996) Bagging predictors. *Mach Learn* 24(2):123–140
- Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. *Neurocomputing* 70(1–3):489–501
- Holland JH (1975) *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. U Michigan Press
- LeCun Y, Boser B, Denker JS et al (1989) Backpropagation applied to handwritten zip code recognition. *Neural Comput* 1(4):541–551
- Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. *Neural Comput* 18(7):1527–1554

38. McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* 5(4):115–133
39. Hochreiter S, Schmidhuber J (1997) Lstm can solve hard long time lag problems. *Advances in neural information processing systems* pp 473–479
40. Smolensky P (1986) *Information processing in dynamical systems: foundations of harmony theory*. Colorado Univ at Boulder Dept of Computer Science, Tech. rep
41. Wan S, Liang Y, Zhang Y, et al (2018) Deep multi-layer perceptron classifier for behavior analysis to estimate parkinson's disease severity using smartphones. *IEEE Access* 6:36,825–36,833
42. Elman JL (1990) Finding structure in time. *Cogn Sci* 14(2):179–211
43. Buitinck L, Louppe G, Blondel M, et al (2013) API design for machine learning software: experiences from the scikit-learn project. In: *ECML PKDD Workshop: languages for data mining and machine learning*, pp 108–122
44. Liu S, McGree J, Ge Z et al (2015) *Computational and statistical methods for analysing big data with applications*. Academic Press
45. Grajski KA, Breiman L, Di Prisco GV, et al (1986) Classification of eeg spatial patterns with a tree-structured methodology: Cart. *IEEE transactions on biomedical engineering BME-33(12):1076–1086*
46. Quinlan JR et al (1996) Bagging, boosting, and c4. 5. *Aaai/iaai* 1:725–730
47. Melville P (2003) *Creating diverse ensemble classifiers*. Computer Science Department, University of Texas at Austin
48. Kumar A (2022) The ultimate guide to adaboost algorithm : What is adaboost algorithm? <https://www.mygreatlearning.com/blog/adaboost-algorithm/>. Accessed 27 March 2022
49. Muthee A (2021) The basics of genetic algorithms in machine learning. <https://www.section.io/engineering-education/the-basics-of-genetic-algorithms-in-ml/>. Accessed 27 March 2022
50. Dastile X, Celik T, Potsane M (2020) Statistical and machine learning models in credit scoring: a systematic literature survey. *Appl Soft Comput* 91(106):263
51. Zhang A, Lipton ZC, Li M, et al (2021) Dive into deep learning. arXiv preprint [arXiv:2106.11342](https://arxiv.org/abs/2106.11342)
52. Masmoudi K, Abid L, Masmoudi A (2019) Credit risk modeling using bayesian network with a latent variable. *Expert Syst Appl* 127:157–166
53. Tran K, Duong T, Ho Q (2016) Credit scoring model: a combination of genetic programming and deep learning. In: *2016 Future Technologies Conference (FTC)*, IEEE, pp 145–149
54. Ha VS, Nguyen HN (2016) Credit scoring with a feature selection approach based deep learning. In: *MATEC Web of Conferences*, EDP Sciences, p 05004
55. Babaev D, Savchenko M, Tuzhilin A, et al (2019) Et-rnn: Applying deep learning to credit loan applications. In: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp 2183–2190
56. Twala B (2010) Multiple classifier application to credit risk assessment. *Expert Syst Appl* 37(4):3326–3336
57. Zhang T, Zhang W, Wei X et al (2018) Multiple instance learning for credit risk assessment with transaction data. *Knowl Based Syst* 161:65–77
58. Roa L, Correa-Bahnsen A, Suarez G et al (2021) Super-app behavioral patterns in credit risk models: financial, statistical and regulatory implications. *Expert Syst Appl* 169(114):486
59. Abakarim Y, Lahby M, Attioui A (2018) Towards an efficient real-time approach to loan credit approval using deep learning. *2018 9th International Symposium on Signal, Image, video and communications (ISIVC)*, IEEE, pp 306–313
60. Dastile X, Celik T (2021) Making deep learning-based predictions for credit scoring explainable. *IEEE Access* 9:50,426–50,440
61. Iwai K, Akiyoshi M, Hamagami T (2020) Structured feature derivation for transfer learning on credit scoring. In: *2020 IEEE International Conference on systems, man, and cybernetics (SMC)*, IEEE, pp 818–823
62. Kumar MR, Gunjan VK (2020) Review of machine learning models for credit scoring analysis. *Ingeniería Solidaria* 16(1)
63. Chi J, Zeng G, Zhong Q, et al (2020) Learning to undersampling for class imbalanced credit risk forecasting. In: *2020 IEEE International Conference on data mining (ICDM)*, IEEE, pp 72–81
64. Arya M, Sastry GH (2020) Deal-'deep ensemble algorithm' framework for credit card fraud detection in real-time data stream with google tensorflow. *Smart Sci* 8(2):71–83
65. Hsu TC, Liou ST, Wang YP et al (2019) Enhanced recurrent neural network for combining static and dynamic features for credit card default prediction. *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp 1572–1576
66. Alam TM, Shaukat K, Hameed IA, et al (2020) An investigation of credit card default prediction in the imbalanced datasets. *IEEE Access* 8:201,173–201,198
67. Yiheng Wei QMYu Qi (2020) Fraud detection by machine learning. *2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, IEEE, pp 101–115
68. Shivanna A, Agrawal DP (2020) Prediction of defaulters using machine learning on azure ml. In: *2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, IEEE, pp 0320–0325
69. Zhang X, Han Y, Xu W et al (2021) Hobo: a novel feature engineering methodology for credit card fraud detection with a deep learning architecture. *Inf Sci* 557:302–316
70. Rushin G, Stancil C, Sun M, et al (2017) Horse race analysis in credit card fraud-deep learning, logistic regression, and gradient boosted tree. In: *2017 systems and information engineering design symposium (SIEDS)*, IEEE, pp 117–121
71. Can B, Yavuz AG, Karsligil EM, et al (2020) A closer look into the characteristics of fraudulent card transactions. *IEEE Access* 8:166,095–166,109
72. Ahmed F, Shamsuddin R (2021) A comparative study of credit card fraud detection using the combination of machine learning techniques with data imbalance solution. In: *2021 2nd International Conference on Computing and Data Science (CDS)*, IEEE, pp 112–118
73. Khatri S, Arora A, Agrawal AP (2020) Supervised machine learning algorithms for credit card fraud detection: a comparison. In: *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, IEEE, pp 680–683
74. Torres RAL, Ladeira M (2020) A proposal for online analysis and identification of fraudulent financial transactions. In: *2020 19th IEEE International Conference on machine learning and applications (ICMLA)*, IEEE, pp 240–245
75. Yu L, Yang Z, Tang L (2016) A novel multistage deep belief network based extreme learning machine ensemble learning paradigm for credit risk assessment. *Flex Serv Manuf J* 28(4):576–592
76. Huang X, Liu X, Ren Y (2018) Enterprise credit risk evaluation based on neural network algorithm. *Cogn Syst Res* 52:317–324
77. Oreski S, Oreski G (2014) Genetic algorithm-based heuristic for feature selection in credit risk assessment. *Expert Syst Appl* 41(4):2052–2064

78. Feng B, Xue W, Xue B, et al (2020) Every corporation owns its image: Corporate credit ratings via convolutional neural networks. In: 2020 IEEE 6th International Conference on Computer and Communications (ICCC), IEEE, pp 1578–1583
79. Dong Y, Xie K, Bohan Z et al (2021) A machine learning model for product fraud detection based on svm. 2021 2nd International Conference on Education. Knowledge and Information Management (ICEKIM), IEEE, pp 385–388
80. Zhou Y, Song X, Zhou M (2021) Supply chain fraud prediction based on xgboost method. 2021 IEEE 2nd International Conference on Big Data. Artificial Intelligence and Internet of Things Engineering (ICBAIE), IEEE, pp 539–542
81. García V, Marqués AI, Sánchez JS (2019) Exploring the synergetic effects of sample types on the performance of ensembles for credit risk and corporate bankruptcy prediction. *Inf Fusion* 47:88–101
82. Giudici P, Hadji-Misheva B, Spelta A (2020) Network based credit risk models. *Qual Eng* 32(2):199–211
83. Chen YR, Leu JS, Huang SA, et al (2021) Predicting default risk on peer-to-peer lending imbalanced datasets. *IEEE Access* 9:73,103–73,109
84. Liang K, He J (2020) Analyzing credit risk among chinese p2p-lending businesses by integrating text-related soft information. *Electron Commer Res Appl* 40(100):947
85. Song Y, Wang Y, Ye X et al (2020) Multi-view ensemble learning based on distance-to-model and adaptive clustering for imbalanced credit risk assessment in p2p lending. *Inf Sci* 525:182–204
86. Niu K, Zhang Z, Liu Y et al (2020) Resampling ensemble model based on data distribution for imbalanced credit risk evaluation in p2p lending. *Inf Sci* 536:120–134
87. Yang J, Li Q, Luo D (2019) Research on p2p credit risk assessment model based on rbm feature extraction-take sme customers as an example. *Open J Busin Manag* 7(4):1553–1563
88. Sirignano J, Sadhwani A, Giesecke K (2016) Deep learning for mortgage risk. arXiv preprint [arXiv:1607.02470](https://arxiv.org/abs/1607.02470)
89. Ostapchenya D (2021) The role of big data in banking : How do modern banks use big data? <https://www.finextra.com/blogposting/20446/the-role-of-big-data-in-banking-how-do-modern-banks-use-big-data>. Accessed 27 March 2022
90. Du G, Liu Z, Lu H (2021) Application of innovative risk early warning mode under big data technology in internet credit financial risk assessment. *J Comput Appl Math* 386(113):260
91. Gao L, Xiao J (2021) Big data credit report in credit risk management of consumer finance. *Wireless Communications and Mobile Computing* 2021
92. Wang H (2021) Credit risk management of consumer finance based on big data. *Mobile Information Systems* 2021
93. Niu A, Cai B, Cai S (2020) Big data analytics for complex credit risk assessment of network lending based on smote algorithm. *Complexity* 2020
94. Pérez-Martín A, Pérez-Torregrosa A, Vaca M (2018) Big data techniques to measure credit banking risk in home equity loans. *J Bus Res* 89:448–454
95. Tang H, Zhang Y, Qiao Q, et al (2020) Risk assessment of credit field based on pso-svm. In: 2020 2nd International Conference on Economic Management and Model Engineering (ICEMME), IEEE, pp 809–813
96. Tomczak JM, Zieba M (2015) Classification restricted boltzmann machine for comprehensible credit scoring model. *Expert Syst Appl* 42(4):1789–1796
97. Lucas Y, Jurgovsky J (2020) Credit card fraud detection using machine learning: A survey. arXiv preprint [arXiv:2010.06479](https://arxiv.org/abs/2010.06479)
98. Wang X, Xu M, Pusuati ÖT (2015) A survey of applying machine learning techniques for credit rating: Existing models and open issues. In: International Conference on neural information processing, Springer, pp 122–132
99. Breeden JL (2020) Survey of machine learning in credit risk. Available at SSRN 3616342
100. Bhatore S, Mohan L, Reddy YR (2020) Machine learning techniques for credit risk evaluation: a systematic literature review. *J Bank Financ Technol* 4(1):111–138
101. Leo M, Sharma S, Maddulety K (2019) Machine learning in banking risk management: a literature review. *Risks* 7(1):29
102. Chi G, Uddin MS, Abedin MZ, et al (2019) Hybrid model for credit risk prediction: an application of neural network approaches. *International Journal on Artificial Intelligence Tools* 28(05):1950,017
103. Najadat H, Altiti O, Aqouleh AA, et al (2020) Credit card fraud detection based on machine and deep learning. In: 2020 11th International Conference on Information and Communication Systems (ICICS), IEEE, pp 204–208
104. Chen X, Li S, Xu X, et al (2020) A novel gsci-based ensemble approach for credit scoring. *IEEE Access* 8:222,449–222,465
105. Wang GG, Deb S, Cui Z (2019) Monarch butterfly optimization. *Neural Comput Appl* 31(7):1995–2014
106. Wang GG, Deb S, Coelho LDS (2018) Earthworm optimisation algorithm: a bio-inspired metaheuristic algorithm for global optimisation problems. *Int J Bioinsp Comput* 12(1):1–22
107. Wang GG, Deb S, Coelho LdS (2015) Elephant herding optimization. In: 2015 3rd international symposium on computational and business intelligence (ISCBI), IEEE, pp 1–5
108. Wang GG (2018) Moth search algorithm: a bio-inspired metaheuristic algorithm for global optimization problems. *Memetic Comput* 10(2):151–164
109. Li S, Chen H, Wang M et al (2020) Slime mould algorithm: a new method for stochastic optimization. *Future Gener Comput Syst* 111:300–323
110. Yang Y, Chen H, Heidari AA et al (2021) Hunger games search: visions, conception, implementation, deep analysis, perspectives, and towards performance shifts. *Expert Syst Appl* 177(114):864
111. Tu J, Chen H, Wang M et al (2021) The colony predation algorithm. *J Bionic Eng* 18(3):674–710
112. Heidari AA, Mirjalili S, Faris H et al (2019) Harris hawks optimization: algorithm and applications. *Future Gener Comput Syst* 97:849–872
113. Ahmadianfar I, Heidari AA, Gandomi AH et al (2021) Run beyond the metaphor: an efficient optimization algorithm based on runge kutta method. *Expert Syst Appl* 181(115):079