[redacted] ([redacted]) [redacted]:[redacted]–[redacted]
https://doi.org/[redacted]/s00[redacted]-[redacted]-[redacted]-[redacted]

REVIEW

# [redacted]-driven credit risk: a systemic review

Si Shi[redacted] · Rita Tse[redacted] · Wuman Luo[redacted] · Stefano D'Addona[redacted] · Giovanni Pau[redacted]

Received: [redacted] February [redacted] / Accepted: [redacted] May [redacted] / Published online: [redacted] July [redacted]


## 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 [redacted] stemmed from a new wave of [redacted] (ML)-driven credit risk models that [redacted]ined tremendous attention from both industry and academia. In this paper, we systematically review a series of major research contributions ([redacted] papers) over the past eight years using statistical, [redacted] 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: [redacted]) most deep learning models outperform classic [redacted] and statistical algorithms in credit risk estimation, and [redacted]) ensemble methods provide higher accuracy compared with single models. Finally, we present summary tables in terms of datasets and proposed models.

[redacted] risk · [redacted] · Deep learning · Statistical learning

## [redacted] Introduction

[redacted] advances heavily affected industry and academia in the past decades, ultimately transforming people's daily life. [redacted] (AI) has been applied to almost every human activity, including pattern recognition, image classification, business, agriculture, transportation, and finance. This paper focuses on [redacted] 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 [redacted]. 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 [[redacted]] and Logistic Regression [[redacted]]. 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 [redacted] and deep learning. Conventional [redacted]

✉ 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@uniroma[redacted].it

[redacted], [redacted], [redacted], [redacted]

[redacted] on [redacted] and [redacted] of [redacted], [redacted], [redacted], [redacted]

[redacted], [redacted], [redacted], [redacted]

[redacted], [redacted], [redacted], [redacted]

[redacted], [redacted], [redacted], [redacted]

techniques, e.g., k-Nearest Neighbor [[redacted]], [redacted] [[redacted]] and [redacted] [[redacted]], are more effective and flexible than statistical methods. In particular, the vital branch of [redacted]-deep learning techniques [[redacted]] 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 [redacted] 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 or[redacted]nized as follows: the survey methodology will be discussed in [redacted]. [redacted]. [redacted]ion [redacted] introduces the principles of statistical learning, [redacted] and deep learning. [redacted]ion [redacted] analyzes credit risk-related applications in detail. In [redacted]. [redacted], presented algorithms are discussed and ranked by their performance a[redacted]inst public datasets. Finally, results and current challenges are summarized in [redacted]. [redacted]; while [redacted]. [redacted] concludes this work.

## [redacted] Survey methodology

### [redacted].[redacted] Methodology

We applied [redacted] (Preferred Reporting Items for Systematic Reviews and Meta-Analyses [redacted]. [redacted]) reviewing methodology in our paper. First, we adopted [redacted] searching platforms for our investi[redacted]tion: [redacted], [redacted], [redacted], [redacted], and [redacted]. We used the keywords "[redacted]" or "deep learning" combined with "credit risk" while searching. We got [redacted] articles in total. Then, we applied a filtering algorithm considering the trade-off between publication year and citations to proceed. After removing [redacted] duplicate records, [redacted] ineligible records, and [redacted] incomplete articles, we obtained [redacted] screened records. Based on the relevance, we excluded [redacted] articles less related to the topic. After manually checking whether the paper has clear evaluation metrics, we further excluded another [redacted] papers. Finally, we kept [redacted] 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.

[redacted]ure[redacted] depicts the [redacted] flow diagram

### [redacted].[redacted] Inclusion and exclusion criteria

In this paper, we select three inclusion criteria: ([redacted]) the relevance of research topic, ([redacted]) the precision of evaluation metrics, ([redacted]) 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 [redacted]. [redacted].

### [redacted].[redacted] 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 [redacted] repository [[redacted]–[redacted]]. In addition, there exist some researches that discover and mine their own data. For example, Chee Kian Leong ([redacted]) uses data from a firm in [redacted] [[redacted]]. Authors in [[redacted]–[redacted]] 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 [redacted] approaches. The traditional [redacted] models for credit risk contain [redacted] (SVMs) [[redacted]], k-Nearest Neighbor (k-NN) [[redacted]], [redacted] (RFs) [[redacted]], Decision Trees (DTs) [[redacted]–[redacted]], AdaBoost [[redacted]], Extreme Gradient Boost (XGBoost) [[redacted]], Stochastic Gradient Boosting ([redacted]) [[redacted]], Bagging [[redacted]], [redacted] ([redacted]) [[redacted]] and [redacted] (Genetic Algorithm) [[redacted]]. Neural network models generally belong to deep learning methods. Most of them include Convolutional [redacted] (CNNs) [[redacted]], Deep Belief [redacted] (DBNs) [[redacted]], [redacted] (ANNs) [[redacted]], LSTM (Long Short-Term Memory) [[redacted]], Restricted Boltzmann Machines (RBMs) [[redacted]], Deep Multi-Layer Perceptron (DMLP) [[redacted]], and [redacted] (RNNs) [[redacted]].

Summary tables and bar charts re[redacted]rding all the methods of the reviewed papers are provided.

### [redacted].[redacted] Taxonomy

The taxonomy is shown as [redacted]. [redacted]. We can divide it into two parts: the first is re[redacted]rding computing technology and the second is credit risk application domain. The two parts are further categorized into sub[redacted]ions. 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.