Annals of Operations Research
https://doi.org/10.1007/s10479-024-05849-1

ORIGINAL RESEARCH

A novel deep learning approach to enhance creditworthiness evaluation and ethical lending practices in the economy

Xiaoyan Qian[1,2] · Helen Huifen Cai[3] · Nisreen Innab[4] · Danni Wang[5] · Tiziana Ciano[6] · Ali Ahmadian[7,8]

Received: 19 September 2023 / Accepted: 16 January 2024

Abstract

Evaluating a borrower's creditworthiness and enabling ethical lending practices are two of the most essential functions of credit scoring, making it an integral part of the economy. Credit risk management is an essential aspect of the financial industry, with the primary goal of minimising potential losses caused by customers failing to meet their credit responsibilities, such as fails to pay and bankruptcies. This risk is inherent in lending activities, where lenders extend credit to individuals or businesses. The traditional credit scoring approaches, which rely on statistical and machine learning techniques to analyse complex data and non-linear correlations in credit data has to be improved. Because the current financial sector lacks credit scoring, a deep learning network-based credit ranking model is presented in this research. This paper applies the complicated field of deep learning known as the stacked unidirectional and bidirectional long short-term memory model in the network to resolve credit scoring issues. Since scoring is not a time sequence issue, the suggested model uses the three-layer stacked LSTM and bidirectional LSTM architecture by modelling public datasets in a new way. Our suggested models beat state-of-the-art, considerably more difficult deep learning methods, proving that we could keep complexity to a minimum. The research findings indicate that the model demonstrates high levels of accuracy across various datasets. The model obtains an accuracy of 99.5% on the Australian dataset, 99.4% on the German dataset (categorical), 99.7% on the German dataset (numerical), 99.2% on the Japanese dataset, and 99.8% on the

✉ Danni Wang
danniwangres1@outlook.com

1 Business School, Nanjing University, Nanjing 210000, China

2 School of Economics and Management, Nanjing Vocational University of Industry Technology, Nanjing 210000, China

3 Middlesex University Business School, London NW4 2BT, UK

4 Department of Computer Science and Information Systems, College of Applied Sciences, AlMaarefa University, P.O. Box 71666, Riyadh 11597, Saudi Arabia

5 School of Continuing Education, Wenzhou Business College, Wenzhou 325035, China

6 Department of Economics and Political Sciences, University of Aosta Valley, Aosta, Italy

7 Decisions Lab, Mediterranea University of Reggio Calabria, Reggio Calabria, Italy

8 Faculty of Engineering and Natural Sciences, Istanbul Okan University, Istanbul, Turkey

Published online: 20 February 2024

Taiwanese dataset. These results highlight the robustness and effectiveness of the model in accurately predicting outcomes for different geographical regions.

Keywords Credit scoring · Deep learning · Long short-term memory (LSTM) · Economy · Credit risk management · Financial

1 Introduction

Financial risk forecasting is paramount in various financial institutions, mainly commercial banks. The process involves the crucial task of credit scoring for each customer. The assessment of creditworthiness provides an integral part in concluding financial institutions when determining whether to approve or deny credit applications from prospective customers (Liu et al., 2023). Credit scoring is typically treated as a categorization assignment in information mining, with the goal being to identify excellent and poor consumers. The banking sector frequently uses credit scores and internal customer ratings to evaluate a client's propensity to repay loans, such as repaying loans and paying interest, as well as meeting other credit conditions. Commercial banks employ these methods to evaluate and classify the risk associated with a customer during the credit application and approval process. The evaluation process recognizes and tracks a customer's credit score changes over time. The study examines the financial and non-financial information of already-established clients during credit and client assessments.

The failure of the Bretton Woods system has had a significant and enduring impact on international financial markets, which have experienced heightened volatility due to the rapid pace of economic globalization (Nikulin & Pekhterev, 2021). In the current financial landscape, it is evident that businesses, financial organizations, and individual investors are confronted with many unforeseen risks (Li & Zhou, 2021). These risks have a detrimental impact on the overall growth and stability of both the national economy and global financial markets (Olivier & Lieven, 2019). The 2008 subprime mortgage crisis is widely recognized for its significant impact on the financial sector, causing a great number of companies to go defunct and major losses for financial organizations (Soares et al., 2021). Moreover, the absence of a comprehensive global financial regulation system exacerbates worldwide financial markets becoming increasingly complicated. In the realm of finance, various entities such as enterprises, insurance companies, and consumers rely on the utilization of robust models for risk forecasting to thoroughly examine financial data and proactively mitigate potential risks (Ouyang et al., 2021). In contrast to manual analysis, the utilization of model prediction offers a more objective approach to interpreting financial data from various perspectives, thereby mitigating the possibility of encountering systematic risks. The viability of the financial market in the country depends largely on the results of an investigation into the system for controlling financial market risks.

Machine learning models have emerged as crucial tools in constructing predictive models. Using classifiers based on machine learning to create credit scores has been the subject of a large body of research. Further investigation is necessary to create the best possible credit score forecasting model. To construct a machine learning model for forecasting that is both robust and accurate, it is crucial to consider the information about the input predictors. Feature selection refers to techniques used to assess features' relevance and significance and reduce the data's dimensionality. There is a great deal of research available to evaluate various feature selection techniques in the context of credit score prediction. These techniques have significantly enhanced the accuracy and effectiveness of credit score prediction models.

For more than 50 years, researchers have examined several categorization algorithms' effectiveness in assessing credit at great length. This is because credit scoring has been at the front of applying ML methods like DTs (Li, 2021), neural networks (NNs) (Wang et al., 2023), and support vector machines (SVMs) (Martens et al., 2007). At first glance, it seemed like these approaches didn't offer significant improvements in accuracy (when compared to the logistic regression model) for determining a person's creditworthiness. However, ensemble approaches, particularly bagging (Jiang, 2023) and boosting (Freund & Schapire, 1997) techniques, have greatly enhanced the efficiency of ML-based scoring techniques. However, despite their widespread use, traditional credit scoring methods may have limitations in capturing complex patterns and non-linear correlations in credit data. As the financial sector faces a lack of credit scoring innovation, there is a compelling need for advanced methods to increase the precision and reliability of credit evaluations.

This paper proposes utilizing deep learning network-based credit scoring models in response to this challenge. Deep learning, a cutting-edge field of artificial intelligence, has witnessed remarkable adoption in various domains but has been relatively underutilized in the credit-scoring industry. Many practical applications, such as image and sound recognition or time series analysis of monetary and economic information, have successfully used DL. In certain settings, DL techniques may train features efficiently, resulting in higher classification results because of the data's temporal and/or geographical correlation. Standard DL models employ data correlations to train feature representations, such as convolutional neural network (CNNs) (Peng & Zhao, 2023) and LSTM (Gao et al., 2023) network. While convolution has been used to discover significant trends in data, one-dimensional (1D) CNN used for data with temporal correlation, such as market guides. To discover important characteristics, existing DL algorithms rely heavily on this learning capability, which captures temporal/spatial correlations (Wang et al., 2022).

This study employs a sophisticated form of deep learning, specifically the BDLSTM and stacked LSTM, to tackle credit score issues. Although credit score is not inherently a timing sequence problem, the authors believe that leveraging the full potential proposed architecture can lead to groundbreaking advancements when applied creatively to credit scoring datasets. Contrary to complex deep learning methods, the suggested credit scoring model aims to balance accuracy and simplicity. By treating and modeling public datasets in a novel manner, the suggested approach is superior to current practices, showcasing its capability to minimize complexity while achieving superior performance. The outcomes of this research open new possibilities for transforming credit risk management and reinforcing ethical lending practices in the financial industry. The successful application of deep learning networks to credit scoring heralds a promising future where innovative technologies contribute to a more robust and efficient global financial ecosystem.

According to Global Credit Market Statistics, the global credit market was valued at approximately USD 4 trillion in 2021, demonstrating the extensive impact of credit scoring systems. According to the Federal Bank in New York, debts of a consumer in the United States will exceed USD 14.6 trillion by 2021. Loan Default Rates In 2020, the average default rate on consumer loans in the United States was reported to be around 2.54%. Loan default rates in emerging markets can be higher; for example, India's gross non-performing assets were around 7.5% in 2021. Financial Setbacks Inadequate Credit Scoring caused a global loss of about USD 2.8 trillion in GDP in 2009 alone due to the 2008 financial crisis, which was partly attributed to poor credit risk assessment.

1.1 Subprime mortgage crisis

Discusses how poor credit scores contributed to the subprime mortgage crisis in the United States, that resulted in over 3.8 million foreclosures. Refer to how JPMorgan Chase used advanced credit scoring models to cut default rates in certain loan categories by 50%. According to a Consumer Financial Protection Bureau (CFPB) study, approximately 26 million Americans, or roughly 11% of the adult population, are "credit invisible," which means they have no credit history with a national consumer reporting agency.

1.2 Comparison with traditional models

Traditional logistic regression models in credit scoring have limitations in dealing with non-linear relationships, with an average accuracy ranging between 60 and 70% in predicting defaults, according to financial industry reports. In some financial institutions' internal studies, deep learning models improved predictive accuracy by up to 20% over traditional models.

2 Related works

During the initial phase of credit scoring method development, the reliance on the personal expertise of professionals was prevalent financial organizations don't have access to a lot of complete historical data, therefore. Over time, the availability of credit data led to the development of numerous methods. In their study, Altman (Xu & Qiu, 2023) developed a Z-score credit scoring model utilizing the principles of multivariate discriminant analysis. This approach was aimed at assessing the creditworthiness of individuals or entities. Furthermore, Parnes (Chen et al., 2022) conducted comprehensive comparative analysis experiments for validating the effectiveness and dominance of the Z-score credit scoring method as proposed by Altman. Logistic regression models have been widely recognized as one of the statistics models closest to the real world. The widespread usage of these methods can be attributed to their notable predictive accuracy, straightforward computational procedures, and robust interpretability, as highlighted in previous research (Luo & Wang, 2023). More and more academic papers are examining for applying machine learning to the field of rating the credits (Bhatia et al., 2017; Xuemin et al., 2023). Traditional machine-learning techniques in this study area fall into two broad buckets: single-classifier techniques and group-learning techniques. Credit scoring has seen substantial study and implementation of several predictors. Notably, decision trees (DTs) (Mandala et al., 2012) and support vector machines (SVM) (Harris, 2015) have been the focus of research and application. Furthermore, a recent study has put forth several enhanced individual classifiers (Abellán & Castellano, 2017).

In their study, Munkhdlai et al. (2021) introduced a novel credit rating methodology that integrates both non-linear and linear techniques. The authors employed a combination of softmax regression, a linear method, and neural network, a non-linear method, to enhance the accuracy and effectiveness of credit scoring. Ensemble learning is a technique that improves the performance of models by constructing and merging base learners. There are two broad classes from which these foundational learners may be constructed: both homogeneously and collaborative learners. When building an ensemble, homogeneity approaches only use one specific kind of base learner. Techniques like random forest (RF) (Liu et al., 2023a) and extreme gradient boosting (XGBOOST) (Yin & Zhang, 2023) are good instances. The goal of diverse ensemble learning is to improve the efficacy of a model by drawing from a wide variety of initial learners. By integrating diverse learning algorithms, this approach aims to exploit each learner's strengths and mitigate their weaknesses. Combining different learners allows for a more comprehensive and robust model that can handle various data patterns and complexities effectively. Through the collaborative efforts of these heterogeneous base learners, the overall recital has ensemble model is expected to surpass that of any single learner. This method has increased significant attention in machine learning and has demonstrated promising consequences in various domains.

Kang et al. (2022) put forth a two-stage credit scoring model in their study. In this study, the initial phase involves credit scoring, which aims to assess the creditworthiness of individuals or entities. This process involves evaluating various factors such as credit history, income, and debt levels to determine the likelihood of timely repayment. Subsequently, the second stage focuses on profit scoring, which evaluates the potential profitability of a particular credit decision. This stage involves analyzing factors such as interest rates, fees, and potential risks to determine the overall financial viability and possible return on investment. Stacked generalization, also known as stacking, was employed in constructing the model. Low-dimensional and carefully built by domain specialists, the data collection characteristics used in these investigations have been noticed.

Recently, the investigation has begun on DL-based credit rating, but it can significantly affect how financial institutions operate. Though, when the number and speed of credit card communications rise, existing methods may struggle to generate strong detection models due to class inequalities and problems with idea drift in credit card fraud data. Sinanc et al. (2021) presented a unique method to handle this problem by combining picture transformation with fraud detection. High-dimensional input data like photographs might be challenging to decipher in the standard CNN architecture. While DL approaches are not yet widely integrated into CNN architectures, a few DL-based classifiers in our review achieved top-five accuracy.

According to research by Liu et al. (2023b), deep learning can allow robots to see, hear, and reason to tackle better-complicated pattern recognition challenges, which in turn helps advance AI technology. Financial applications of deep learning have reached a mature stage. Combining LSTM with RNN for price extrapolation or market movement classification, using a CNN for factor mining, Guo et al. (2019) constructed a stock selection model. The deep learning method may save training time, boost network performance, and enhance generalizability by automatically tuning the model parameters. The Ensemble model combines a Support Vector Regression (SVR) machine, an extreme value optimizing method, and a neural network (NN) for nonlinear neural networks for time series forecasting by Qian et al. (2020). A financial sector risk administration strategy based on AI was suggested and built by Sha et al. (2020). The system's economic portfolio was well-suited to the preexisting framework for managing risks in the financial markets, and it offered high practical value (Sha et al., 2020). Utilizing the metaheuristic approach and the deep learning model, Bachute et al. (2021) built a novel biological heuristic meta-start technique to actualize spatiotemporal forecasting of financial goods.

The paper (Alipour & Bastani, 2305) explores Value-at-Risk-based portfolio insurance, comparing its performance with Constant Proportion Portfolio Insurance (CPPI) in a market characterized by Markov-modulated regime switching. This study is significant for its nuanced analysis of portfolio risk management strategies in complex market conditions, offering valuable insights for investors and financial analysts. The research (Razmi et al., 2022) discusses about effects of oil prices indirectly on consumer spending through assets, published in the International Journal of Energy Economic and Policies. The study's relevance lies in its examination of the intricate relationship between commodity prices and economic consumption patterns, providing a broader understanding of economic dynamics in the context of fluctuating oil prices.

Focusing on the Tehran Stock Exchange, this study (Rostami et al., 1999) analyzes the relationship among trading volume, return volatility and stock returns. Despite its earlier publication date, the research contributes to the understanding of market dynamics in emerging economies, offering a historical perspective on financial market behaviors. The article (Alahdadi et al., 2023) introduces a novel double processing model for cloud computing in resource allocation, aiming for truthfulness and budget balance. The model's relevance is underscored by the growing importance of efficient resource allocation mechanisms in cloud computing environments, a key concern for both cloud service providers and users.

The study (Tehranian & Can, 2308) investigates the potential of machine learning in predicting economic declines using market sentiments. This research is timely and innovative, considering the increasing reliance on machine learning for economic forecasting and the need for more sophisticated models to interpret complex market indicators. The paper (Dehghani & Larijani, 2023) proposes an algorithm to predict stock market indices using a combination of the MID algorithm and neural networks. The research is particularly relevant for its application of neural networks in financial forecasting, demonstrating the growing intersection of machine learning and financial analysis.

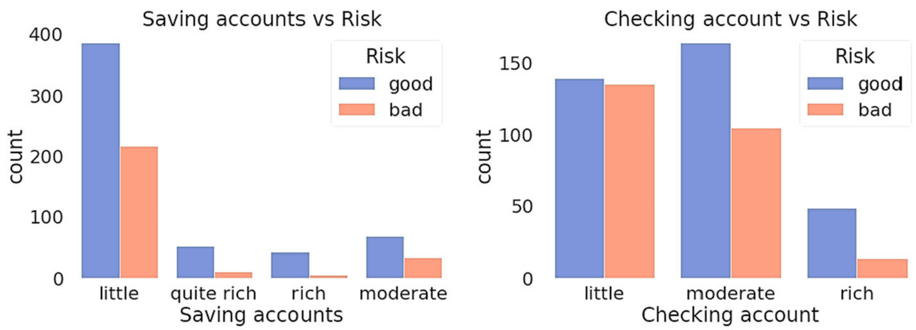
3 Methods and materials

The research purposes is to present and analyse a deep learning-based credit scoring model that can address some of the shortcomings of more traditional credit scoring methods. The use of stacked based unidirectional-bidirectional LSTM networks, specifically long short-term memory, will be employed in the model to effectively capture intricate patterns and non-linear correlations in credit data. The suggested model will be analyzed and compared to this study's most recent advances in deep learning techniques. Examining how well the model suggested works is the focus of this study in improving credit risk management and facilitating ethical lending practices. The collection of publicly available credit datasets will be undertaken for training and evaluating the model. The datasets should encompass various credit-related attributes, encompassing historical borrower information and the corresponding credit risk classifications.

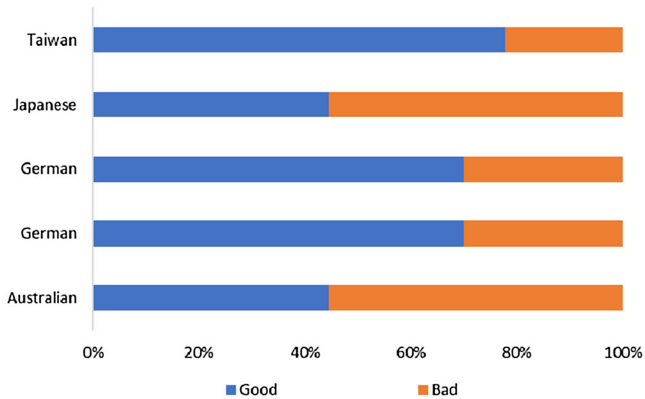
3.1 Data collection

The predictive model uses different datasets on credit scoring. In this study, we assess the presentation of deep learning model classifiers on various credit datasets, including the Australian, German (numerical), German (categorical), Taiwanese and Japanese datasets. The Australian Credit Approval Dataset comprises information about credit card requests. The dataset consists of 690 instances and 14 attributes, one of them being the class attribute representing the credit approval status (label). The dataset contains 6 numerical attributes and 8 categorical attributes. To ensure the privacy of the information, however, the designations of attributes and corresponding values have been modified and changed. Additionally, the original labels of the categorical attributes have been changed to numerical for the expediency of statistical algorithms. This diverse mix of attributes and the careful transformation make the Australian Credit Approval Dataset a valuable resource for exploring creditworthiness and enhancing credit risk management practices. The information set is intriguing because of its variety of attribute types. There are constant and conventional properties and few and many values for each. These attribute types can provide valuable insights for credit scoring and risk assessment models. Furthermore, it's worth noting that the dataset may have some missing values, which would require data preprocessing and handling techniques before using the data for modeling or analysis.

The Fig. 1a shows the Risk distribution from the saving and checking accounts from the German Credit Risk Analysis dataset and Fig. 1b shows the number of bad and good score from overall datasets. The German Credit dataset consists of two sets of data, the original dataset with 1000 instances and 20 attributes, comprising 7 numerical and 13 categorical attributes. To create the "german.data-numeric" file, the categorical attributes in the original dataset were transformed into numerical representations. This was achieved by introducing indicator variables to accommodate algorithms that cannot handle categorical variables directly. Additionally, attributes originally ordered categorical, such as attribute 17, were encoded as integers. With the introduction of the "german. data-numeric" file, all 24 attributes are now represented numerically, making it suitable for algorithms that rely on



(a)



(b)

Fig. 1 **a** Risk distribution from the saving and checking accounts from the German Credit Risk Analysis dataset **b** Number of bad and good score from overall datasets numerical data input. The dataset can now be utilized effectively with various statistical and machine learning algorithms to analyze credit-related patterns and assess credit risk.

3.2 Pre-processing

To guarantee the data is usable for modeling and examination, preparation is an essential stage in statistical analysis and machine learning, which is also true when interacting with the German Credit dataset. The first step in data analysis involves checking for missing values within the dataset. There are several potential causes of missing numbers, including typographical mistakes during data entry, equipment failure, or unresponsive participants. Finding and handling these gaps in data properly is critical for analytical reliability and validity. To check for missing values, one can examine each variable in the dataset and determine if any observations are incomplete or contain null values. This can be done by visually inspecting the dataset or using statistical functions to identify missing values. Additionally, it is important to consider any specific coding conventions used to represent missing values, such as "NA". Once missing values are identified, the next step Common methods for handling missing values in research studies typically involve removing instances with missing values or imputing the missing values using measures such as mean, median, or mode.

Since the original dataset contains categorical attributes, they need to be encoded into numerical values for most machine learning algorithms to process them. Multiple-category can be signified as binary vectors using the one-hot encoding. The binary representation of an assortment is a vector in which everything saves the element representing the grouping has been set to one. Suppose we have a categorical attribute "A" with 'n' categories(A_1, A_2, ..., A_n). The one-hot encoding for a data point $'x'$ with category $A = A_i$ is given by:

$$OneHot(x, A) = [0, 0, ..., 1(atindexi), ..., 0] \tag{1}$$

Z-score normalization, also known as standardization, is a common data preprocessing technique used to scale numerical data to have zero mean and unit adjustment. This procedure ensures that all the numerical features are on the same scale, making them comparable and preventing features with larger values from dominating the learning process in certain algorithms. The Z-score normalization formula for a data point 'x' with respect to a feature $'X'$ is as follows:

$$Z - score(x, X) = \frac{(x - mean(X))}{standard\_deviation(X)} \tag{2}$$

where the original value is x of the data point for feature$X$. $mean(X)$, mean (average) data of all data points for feature$X$. $standard\_deviation(X)$ is the deviation of all data points for feature$X$.

When there is an inequity across classes in a dataset, data enhancement methods like SMOTE (Synthetic Minority Over-sampling Technique) can help. It generates minority class populations by integrating among real minority class populations.

Algorithm 1 Synthetic minority over-sampling technique

1. Select a minority class instance (xi) from the dataset.
2. Find its k nearest neighbor (N) belonging to the same minority class.
3. Randomly choose one of the k neighbors (xj).
4. Generate a synthetic data point (xi_{new}) by combining xi and xj:
   a. $$xi_{new} = xi + rand(0,1) * (xj - xi) \quad (3)$$

   where $rand(0,1)$ is a random number between 0 and 1.
5. Repeat steps 1 to 4 for the desired number of synthetic samples to be generated.

SMOTE effectively rises the minority class instances, balancing the class distribution and mitigating the issue of imbalance. Algorithms for machine learning, particularly ones that are sensitive to unbalanced data, may operate better with the aid of SMOTE's synthetic samples since they are less likely to be skewed towards the overwhelming class.

3.3 Methodology

In the realm of sequence modelling tasks, such as credit scoring, the utilization of a three-layer stacked bidirectional LSTM network has been proposed. This network architecture consists of two initial layers that employ bidirectional LSTMs, while the final layer employs a unidirectional LSTM. The motivation behind this design is to effectively capture long-range dependencies and non-linear patterns, which are of utmost importance in credit scoring and similar tasks. The architecture under consideration contains the utilization of bidirectional LSTM models in the initial two layers. These bidirectional LSTMs are capable of processing the input order in both the forward and backward instructions. Subsequently, the final layer incorporates a unidirectional LSTM that leverages the bidirectional output to generate the ultimate prediction.

The initial set of features is fed into the first layer of the bidirectional LSTM. The initial layer has two sub-layer of bidirectional LSTM units. Both the moving forward and backward directions of the input sequence are processed simultaneously by these sub-layers. Combining the results of two bilateral LSTM units yields this layer's outcome. The second layer is composed of two sub-layers of bidirectional LSTM units. The first layer's output series is split in half and processed forward and backward by two sub-layers. The outcome of this layer is calculated by summing the results of the two bidirectional LSTM models, as has been done in earlier studies. The Fig. 2 likely illustrates the architecture of a deep learning model designed to work with complex patterns and dependencies in sequential data.

The third and ultimate layer consists of a unidirectional LSTM model, which receives the output sequence from the second layer and performs computations in a single direction. The output generated by this layer will be utilized in the final prediction process. In the context of neural networks, the output layer is responsible for transforming the final hidden state of the third layer, which corresponds to the output of the unidirectional Long Short-Term Memory (LSTM) layer. This transformation involves mapping the hidden state to the prediction space. In the context of binary classification tasks such as credit scoring, it is common to use a activation function called sigmoid in the output node. This activation function allows for the calculation of a probability score. One of the primary benefits associated with the utilization of a stacked bidirectional LSTM architecture is its ability to effectively capture intricate and extensive dependencies within the input sequence. The bidirectional nature of the model allows for the incorporation of future as well as past context, which is beneficial for capturing a

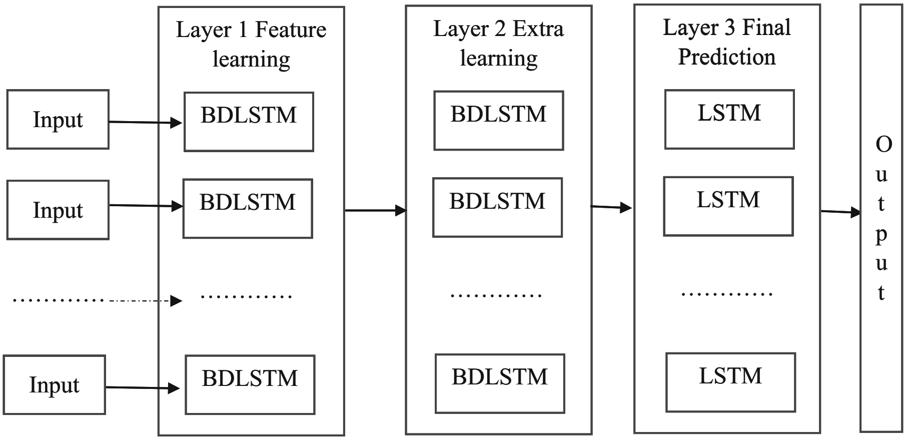


Fig. 2 Proposed Stacked LSTM and Bi-Directional LSTM

more comprehensive understanding of the input. Additionally, the stacking of multiple layers in the model further improves its capacity to learn hierarchical features, enabling it to extract and represent complex patterns in the data.

3.3.1 LSTM architecture

The LSTM architecture is a type of RNN considered to model and process sequential data with long-range dependencies. Unlike traditional RNNs, LSTM networks can effectively overcome the vanishing gradient problem, enabling them to capture and retain information over longer time intervals. The core idea behind LSTM is the introduction of memory cells, also known as memory blocks or cells, which are responsible for storing and updating information over time.

The cell state serves as the memory of the LSTM. It is the long-term memory component that allows the model to retain relevant information over time. The cell state is updated through various operations involving gates, allowing it to remember or forget certain information. The present input and the prior concealed state are both taken into account when deciding which data from each is to be utilized to change the cell's state. It regulates how much new data is stored in long-term memory. The forget gate selects the memories from the prior state of the cell to be remembered. It aids the LSTM's ability to remember the most important information while discarding the rest. Figure 3 depicts the LSTM framework.

The LSTM architecture can be signified with the subsequent equivalences for each time step 't':

The Input Gate (i_t),

$$i_t = sigmoid(W_i * [h_{(t-1)}, x_t] + b_i) \tag{4}$$

$$sigmoid(y) = \frac{1}{1 + e^{-z}} \tag{5}$$

The Forget Gate (f_t),

$$f_t = sigmoid(W_f * [h_{(t-1)}, x_t] + b_f) \tag{6}$$

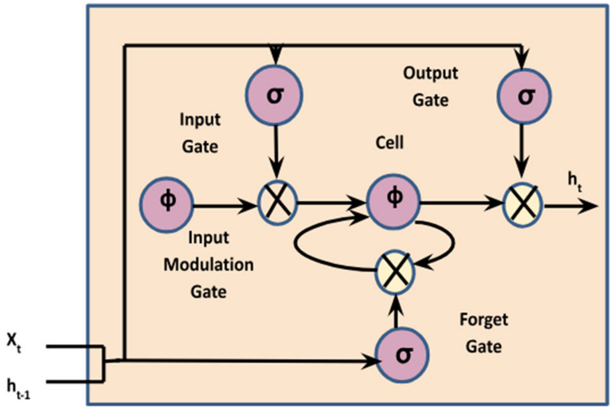


Fig. 3 LSTM architecture

The Candidate Cell State ($\widetilde{C}_t$),

$$\widetilde{C}_t = \tanh(W_c * [h_{(t-1)}, x_t] + b_c) \tag{7}$$

$$\tanh(y) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \tag{8}$$

The Cell State Update (C_t),

$$C_t = f_t * C_{(t-1)} + i_t * \widetilde{C}_t \tag{9}$$

The Output Gate (o_t),

$$O_t = sigmoid(W_o * [h_{(t-1)}, x_t] + b_o) \tag{10}$$

The Hidden State (h_t),

$$h_t = O_t * \tanh(C_t) \tag{11}$$

where x_t is the time step of input t. $h_{(t-1)}$ is the previous times hidden state step (initially set to 0 for the first-time step). W_i, W_f, W_c, W_o are the weight matrices for corresponding gate, respectively. b_i, b_f, b_c, b_o are the bias vectors for gates. The LSTM architecture's capability to capture long-range dependencies and remember information for extended periods makes it well-suited for various order task modelling, such as time series forecasting, and credit scoring, where data exhibits temporal dependencies.

3.3.2 Bi-Directional mode of LSTM architecture

The Bi-Directional LSTM (BLSTM) architecture is an extension of the standard LSTM architecture that enhances the model's ability to record historical and future contexts of a set of events. Unlike the unidirectional LSTM, which processes the input sequence sequentially from one direction (e.g., past to future or future to past), the BLSTM performs forward and reverse processing of the order concurrently. This bidirectional operation enables the BLSTM to capture context from both ends of the sequence, making it well-suited for tasks that require a comprehensive understanding of the input data.

The BLSTM is a neural network construction that is designed to develop input orders in a bidirectional manner. The forward LSTM model sequentially analyses the input sequence starting from the initial time step and progressing towards the final time step, following a past-to-future direction. The computation of hidden states and cell states is performed iteratively as the model progresses through the sequence. The backward LSTM is a special case of the LSTM model that processes the order of inputs backwards, from the most recent time step to the earliest. Incorporating potential data into the model's interpretation at the current time step is made possible by using this method. By considering the sequence in a backward manner, the backward LSTM can potentially capture different patterns and dependencies compared to its forward counterpart. In addition, the computation of hidden states and cell states is performed. In the context of sequential data processing, it is common to employ BDLSTM networks. Hidden levels produced by the front LSTM and the reverse LSTM are concatenated at every increment.

This concatenation operation results in the formation of the final hidden state of the BDLSTM. The input to the output layer for making forecasts or further processing in various tasks involves utilizing the final hidden states obtained from the concatenated forward and backward LSTMs. The Fig. 4 shows the stacked BDLSTM architecture.

The formation of the final hidden state at each time step is achieved by concatenating the hidden states from both the forward and backward directions. The Bi-Directional LSTM architecture has been found to be particularly useful in tasks that require a comprehensive understanding of context from both previous and upcoming time steps. The BDLSTM's skill to effectively record long-range relationships and improve its capacity to discover complicated patterns within the input data is enhanced by the inclusion of knowledge from both perspectives. This tool's flexibility makes it useful for a broad variety of NLP and model building applications.

Here are the formulas for the forward and backward computations in a Bi-Directional LSTM (BLSTM) architecture for a given time step 't':

The Forward LSTM equations:

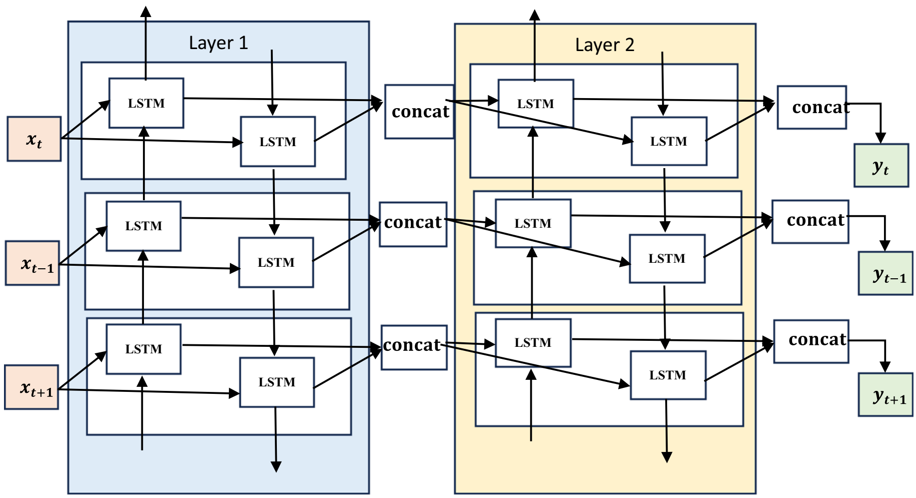


Fig. 4 The stacked BDLSTM architecture

The Input Gate (i_t),

$$i_t = \sigma(W_{\{ix\}} * x_t + W_{\{ih\}} * h_{\{t-1\}} + b_i \tag{12}$$

The Forget Gate (f_t),

$$f_t = \sigma(W_{\{fx\}} * x_t + W_{\{fh\}} * h_{\{t-1\}} + b_f) \tag{13}$$

The Candidate Cell State (g_t)

$$g_t = \tanh\left(W_{\{gx\}} * x_t + W_{\{gh\}} * h_{\{t-1\}} + b_g\right) \tag{14}$$

The Cell State (c_t):

$$c_t = \left\{f_t * c_{\{t-1\}}\right\} + i_t * g_t \tag{15}$$

Output Gate (o_t):

$$o_t = \sigma(W_{\{ox\}} * x_t + W_{\{oh\}} * h_{\{t-1\}} + b_o) \tag{16}$$

Hidden State (h_t):

$$h_t = o_t * \tanh(c_t) \tag{17}$$

where x_t is time of input t. $h_{\{t-1\}}$ is the previous time steps hidden state (initially set to 0 for the first-time step). $c_{\{t-1\}}$ is the previous time step cell state (initially set to 0 for the first-time step). σ represents a sigmoid activation symbol. Tanh is the tangent hyperbolic based activation function.

$W_{\{ix\}}$, $W_{\{ih\}}$, $W_{\{fx\}}$, $W_{\{fh\}}$, $W_{\{gx\}}$, $W_{\{gh\}}$, $W_{\{ox\}}$, $W_{\{oh\}}$ are the Weight matrices. b_i, b_f, b_g, b_o are the bias vectors. The Backward LSTM equations:

The backward LSTM uses similar equations as the forward LSTM but with different weight matrices and bias vectors:

The Input Gate ($i\prime t$),

$$i\prime_t = \sigma\left(W\prime_{\{ix\}} * x_t + W\prime_{\{ih\}} * h\prime_{\{t-1\}} + b\prime_i\right) \tag{18}$$

The Forget Gate ($f\prime t$),

$$f\prime_t = \sigma\left(W\prime_{\{fx\}} * x_t + W\prime_{\{fh\}} * h\prime_{\{t-1\}} + b\prime_f\right) \tag{19}$$

The Candidate Cell State ($g\prime t$),

$$g\prime_t = \tanh\left(W\prime_{\{gx\}} * x_t + W\prime_{\{gh\}} * h\prime_{\{t-1\}} + b\prime_g\right) \tag{20}$$

The Cell State ($c\prime_t$),

$$c\prime_t = f\prime t * c\prime_{\{t+1\}} + i\prime t * g\prime t \tag{21}$$

The Output Gate ($o\prime t$),

$$o\prime_t = \sigma\left(W\prime_{\{ox\}} * x_t + W\prime_{\{oh\}} * h\prime_{\{t-1\}} + b\prime_o\right) \tag{22}$$

The Hidden State ($h\prime_t$),

$$h\prime_t = o\prime_t * \tanh(c\prime_t) \tag{23}$$

where x_t is time of input step t. $h\prime_{\{t-1\}}$ is the backward LSTM hidden state from the next time step. $c\prime_{\{t+1\}}$ is the backward LSTM based cell state from the next time step. $W\prime_{\{ix\}}$, $W\prime_{\{ih\}}$, $W\prime_{\{fx\}}$, $W\prime_{\{fh\}}$, $W\prime_{\{gx\}}\}$, $W\prime_{\{gh\}}$, $W\prime_{\{ox\}}$, $W\prime_{\{oh\}}$ are the Weight matrices for the gates. $b\prime_i$,

$b\prime_f, b\prime_g, b\prime_o$ are the Bias vectors for the backward forget gate, input gate, candidate cell state, and output gate, respectively. Combining the hidden states from the previous and next time steps yields the final concealed state at each time step,

$$h_t = [h_t, h\prime_t] \tag{24}$$

4 Result and discussion

The neural networks were implemented in Keras library with python. Keras is a neural network API that is implemented in Python and has the ability to be executed on either TensorFlow or Theano frameworks. The networks were trained on a computer equipped with an Intel Core i5-3320 M CPU and 8 GB of RAM. In this study, the training duration for the neural networks on the central processing unit (CPU) ranged from 10 min to 4 h. In order to assess the effectiveness of our network, a comprehensive evaluation is conducted by considering four key metrics: Accuracy, Precision, Recall, and F1 are mentioned in Table 1. These metrics are utilized to provide a thorough analysis of the network's performance. In this investigation, the dataset was randomly split into a set for training and a test set. In particular, the training set comprised 80% of the data, while the test set constituted 20%.

Table 1 Architecture of the proposed model

Layer #	Layer type	Number of units/nodes	Activation function	Direction	Notes
1	Input Layer	14	–	–	Dimensionality based on input features
2	Forward LSTM Layer 1	64	tanh	Forward	Processes input sequence from start to end
3	Backward LSTM Layer 1	64	tanh	Backward	Processes input sequence from end to start
4	Concatenation Layer	–	–	Bidirectional	Concatenates outputs of forward and backward LSTM layers
5	Forward LSTM Layer 2	128	tanh	Forward	Further processing of concatenated sequence
6	Backward LSTM Layer 2	128	tanh	Backward	Further processing of concatenated sequence
7	Concatenation Layer	–	–	Bidirectional	Concatenates outputs of second forward and backward LSTMs
8	Dense Layer	32	ReLU	–	Additional processing and feature extraction
9	Output Layer	14	softmax/sigmoid	–	Adjusted based on the specific task (classification/regression)

Furthermore, within the training set, a random selection technique was utilized to extract 10% of the data for the purpose of parameter verification in the deep learning models.

In this study, the accuracy rates of various datasets are depicted in Fig. 5. Each dataset corresponds to a unique category or nationality. The accuracy achieved by the Australian dataset was found to be 99.5%, suggesting a high level of effectiveness in accurately classifying data pertaining to Australia. In a similar vein, the dataset pertaining to German information, which falls under the categorical domain, achieved an impressive accuracy rate of 99.4%. This outcome serves as a testament to the model's exceptional capability to effectively categorize and classify German-related data. Table 2 shows the performance of various deep learning models in credit scoring system.

Additionally, the German dataset (numerical) demonstrated even greater accuracy, reaching an impressive 99.7%, highlighting the model's precision in handling numerical data from Germany. The Japanese dataset yielded a commendable accuracy of 99.2%, suggesting the model's strong capability in effectively processing and classifying Japanese-related data. Lastly, the Taiwanese dataset exhibited exceptional accuracy, recording an astounding 99.8%, reinforcing the model's exceptional performance in accurately categorizing data associated with Taiwan. These high accuracy rates across various datasets indicate the model's robustness and competence in handling diverse data types and nationalities.

In Fig. 6, the loss values for credit scoring on different datasets are presented. A greater accuracy is shown by a smaller loss, which is a measure of how much the forecasts made by the model deviate from the actual outcomes. For the Australian dataset, the loss is impressively low at 0.03, suggesting that the credit scoring model performs very well in accurately predicting creditworthiness for Australian individuals. Similarly, for the German dataset with categorical features, the loss is 0.04, indicating strong performance in credit scoring for German applicants based on categorical data. The German dataset with numerical features achieves an even lower loss of 0.02, indicating the model's exceptional ability to predict credit scores for German applicants using numerical information. On the other hand, the Japanese dataset has a slightly higher loss of 0.08, suggesting that the model faces a bit more difficulty in accurately scoring creditworthiness for Japanese individuals. Finally, the Taiwanese dataset demonstrates the best performance, with an extremely low loss of 0.01, showcasing the model's outstanding accuracy in credit scoring for Taiwanese applicants. Overall, these loss values provide valuable insights into the model's effectiveness in credit scoring across different datasets, with particularly outstanding performance on the Taiwanese dataset.

Table 3 and Fig. 7 presents a comprehensive performance comparison with existing literature surveys on credit scoring methodologies for various datasets. The table includes references to different research works, the datasets used, the applied methodologies, and the corresponding accuracy achieved by each method. For the Australian dataset, Kuppili et al. (2020) utilized the Extreme Learning method, achieving an accuracy of 95.5%. Radović et al. (2021) employed Ensemble classifiers and obtained an accuracy of 92%, while Zhang et al. (2021) applied Ensemble learning to achieve a slightly higher accuracy of 92.3%. Moving on to the German dataset with categorical features, Trivedi (2020) employed an ML classifier and attained an accuracy of 93.1%. Arora et al. (2020) used the Bolasso feature selection method and achieved an accuracy of 84.0%. On the other hand, Zhang et al. (2020) utilized an Ensemble model and achieved an accuracy of 68.4% (Table 4).

For the German dataset with numerical features, Tripathi et al. (2021) used a DNN (Deep Neural Network) and achieved an accuracy of 88.2%. Wang et al. (2018) utilized the Genetic Algorithm (GA) method, resulting in an accuracy of 78.5%. Liu et al. (2022) applied Boosting DT (Decision Trees) and attained an accuracy of 77.15%. For the Japanese dataset, Zhang et al. (2021) employed an Ensemble model and realized an accuracy of 93.1%. He et al. (2023)

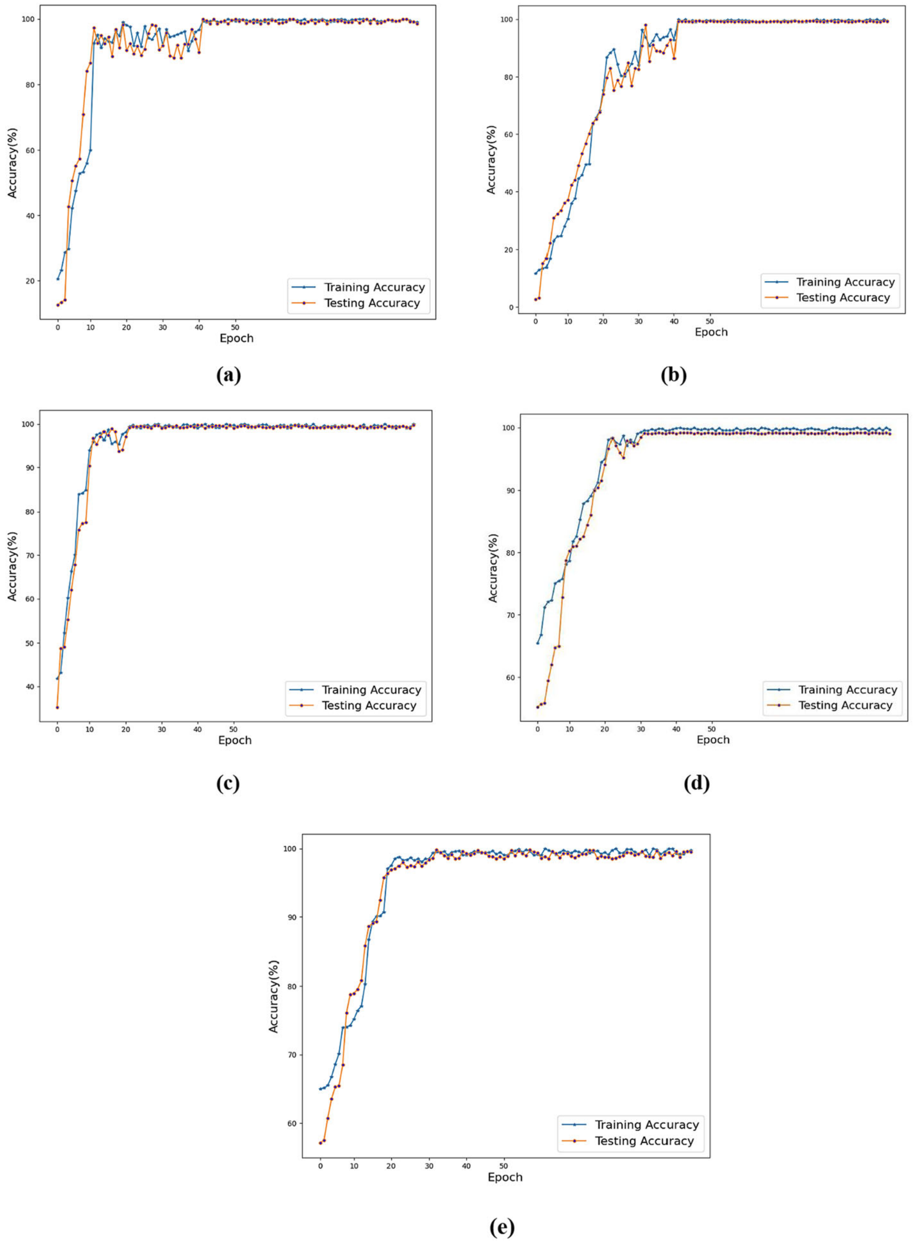


Fig. 5 Accuracy of credit scoring on the **a** Australian = 99.5% 99.8% **b** German (categorical) = 99.4% **c** German (numerical) = 99.7% **d** Japanese **e** Taiwanese = 99.2%

Table 2 Performance Analysis of credit scoring on five datasets

Dataset	Accuracy (%)	Precision	Recall	F1-score
Australian	99.5	0.99	0.98	0.99
German (categorical)	99.4	0.99	0.99	0.99
German (numerical)	99.7	1.00	1.00	0.99
Japanese	99.2	0.99	0.99	1.00
Taiwanese	99.8	1.00	0.99	0.99

utilized TDNN (Time-Delay Neural Network) and achieved an accuracy of 90.4%, while Tripathi et al. (2020) employed ELM (Extreme Learning Machine) and reached an accuracy of 94.6%. Lastly, for the Taiwanese dataset, Liu et al. (2022) used DT (Decision Trees) and achieved an accuracy of 93.6%. Li et al. (2021) utilized an Ensemble Model and achieved an accuracy of 82.7%. Li et al. (2020) applied KNN (K-Nearest Neighbors) and achieved an accuracy of 89.4%. The table provides a comprehensive overview of various credit scoring methodologies and their respective accuracies for different datasets, aiding researchers and practitioners in choosing the most appropriate approach for scoring the credits in various scenarios.

Cross-validation is essential for determining the model's generalizability and robustness across different datasets. The datasets (Australian, German, Japanese, and Taiwanese) are subdivided into 'k' subsets. The model is trained on 'k − 1' subsets before being validated on the remaining subset. This process is repeated 'k' times, with each subset used for validation exactly once. Key performance metrics such as accuracy, precision, recall, and F1-score should be calculated for each fold of cross-validation. This aids in understanding the model's consistency across different data segments. The model's overall effectiveness and reliability are summarized by the average of the performance metrics across all folds (Table 5).

5 Conclusion and future work

In conclusion, credit scoring plays a crucial role in evaluating a borrower's creditworthiness and promoting ethical lending practices, making it an indispensable aspect of the economy and financial sector. Effective credit risk management heavily relies on accurate credit scoring to mitigate the risk of customer defaults and bankruptcies. The objectives encompass improved financial market management, comprehensive individual credit evaluations, and reduced business risk within the financial industry. While traditional techniques have been widely used, they might have limitations in capturing complex patterns and non-linear correlations present in credit data. Recognizing the current challenges faced by the financial sector, this research proposes the adoption of deep learning network-based credit scoring models. Despite the increasing popularity of deep learning in various fields, its application in the credit scoring industry remains limited. This study employs a sophisticated deep learning approach known as stacked unidirectional and bidirectional LSTM networks to address credit scoring challenges. Despite credit scoring not being inherently a time sequence problem, the proposed robust model harnesses the architecture by creatively modeling public datasets in a novel manner. The suggested models outperform state-of-the-art and more complex deep learning methods, proving that the research team successfully managed to strike a balance

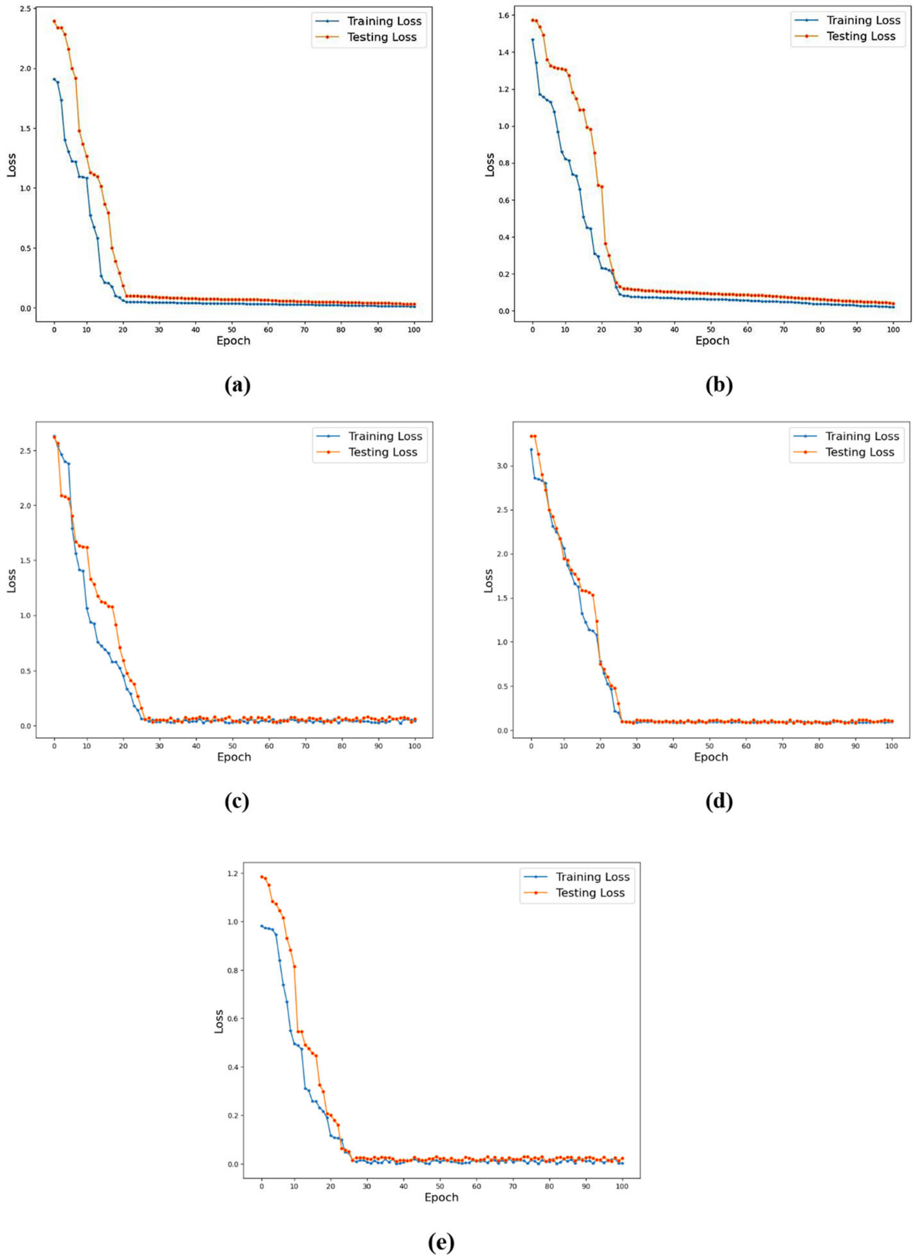


Fig. 6 Loss of credit scoring on the **a** Australian = 0.03 **b** German (categorical) = 0.04 **c** German (numerical) = 0.02 **d** Japanese = 0.08 **e** Taiwanese = 0.01

Table 3 The comparison of various deep learning performance

Method	Precision	Recall	F1-score
Genetic algorithm (Wang et al., 2018)	0.93	0.92	0.93
XGBOOST(Liu et al., 2022)	0.85	0.86	0.86
LSTM (Ouyang et al., 2021)	0.97	0.96	0.97
Proposed BDLSTM	0.98	0.99	1.00

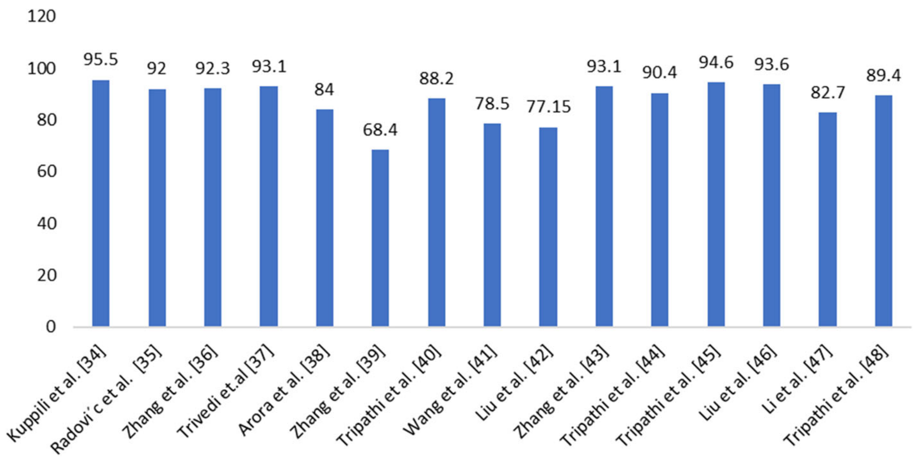


Fig. 7 Performance Comparison with existing literature survey

Table 4 Performance Comparison with existing literature survey

References	Dataset	Methodology	Accuracy
Kuppili et al. (2020)	Australian	Extreme Learning	95.5
Radović et al. (2021)	Australian	Ensemble classifiers	92
Zhang et al. (2021)	Australian	Ensemble learning	92.3
Trivedi (2020)	German (categorical)	ML classifier	93.1
Arora et al. (2020)	German (categorical)	Bolasso feature selection	84.0
Zhang et al. (2020)	German (categorical)	Ensemble model	68.4
Tripathi et al. (2021)	German (numerical)	DNN	88.2
Wang et al. (2018)	German (numerical)	GA	78.5
Liu et al. (2022)	German (numerical)	Boosting DT	77.15
Zhang et al. (2021)	Japanese	Ensemble model	93.1
He et al. (2023)	Japanese	TDNN	90.4
Tripathi et al. (2020)	Japanese	ELM	94.6
Liu et al. (2022)	Taiwanese	DT	93.6
Li et al. (2021)	Taiwanese	Ensemble Model	82.7
Li et al. (2020)	Taiwanese	KNN	89.4

Table 5 Cross validation results in real time

Fold	Accuracy (%)	Precision	Recall	F1-score
1	98.7	0.99	0.98	0.98
2	99.1	0.98	0.99	0.99
3	98.9	0.97	0.99	0.98
4	99.3	0.99	0.98	0.99
5	98.8	0.98	0.97	0.97
Average	99.0	0.98	0.98	0.98

between model simplicity and superior performance. The model achieves accuracy on Australian99.5%, German (categorical)99.4%, German (numerical) 99.7%, Japanese 99.2%, and Taiwanese 99.8%. For future work, further exploration and research into deep learning-based credit scoring models are recommended. Investigating other variations of LSTM architectures, incorporating additional features, and experimenting with more diverse datasets could yield even more promising results. Additionally, efforts to enhance model explainability and interpretability could facilitate better decision-making in credit risk management. Moreover, evaluating the proposed models in real-world scenarios and conducting comparative studies with traditional methods would provide valuable insights into their practical applicability and overall effectiveness in the credit scoring industry.

The limitation of model's are reliance on specific datasets, as well as potential biases in historical data, may limit its applicability across diverse demographic and economic contexts. Future research should concentrate on improving the model's adaptability to different socioeconomic environments, improving its interpretability, and incorporating a broader range of socioeconomic factors to improve its predictive accuracy. Real-world applications and comparative studies with traditional credit scoring methods will be critical in validating the model's practical effectiveness and identifying areas for further improvement.

Funding National Natural Science Foundation of China (General Program) (72372073) The University Philosophy and Social Science Major Fund Project in Jiangsu (2023SJZD061) Start-up Fund for Talented Researchers of Nanjing Vocational University of Industry Technology (2022SKYJ03). Besides, Nisreen Innab, would like to express sincere gratitute to AlMaarefa University, Riyadh, Saudi Arabia, for supporting this research. Authors also would like to appreciate the University Philosophy and Social Science Major Fund Project in Jiangsu Province (2023SJZD061).

Data availability The datasets generated during and/or analysed during the current study are available in the [kaggle] repository, German: [http://archive.ics.uci.edu/ml/datasets/Statlog+%28German+Credit+Data%29]. or Kaggle url: https://www.kaggle.com/uciml/german-credit. Australian: [http://archive.ics.uci.edu/ml/datasets/Statlog+%28Australian+Credit+Approval%29]. Japan: [http://archive.ics.uci.edu/ml/datasets/Japanese+Credit+Screening]. Taiwan: [http://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients]. or Kaggle [https://www.kaggle.com/uciml/default-of-credit-card-clients-dataset].

Declarations

Conflict of interest The authors declare that they have no competing interests.

Ethical approval and consent to participate Not applicable.

Consent for publication Not applicable.

References

Abellán, J., & Castellano, J. G. (2017). A comparative study on base classifiers in ensemble methods for credit scoring. *Expert Systems with Applications, 73*, 1–10.

Alahdadi, A., Safaei, A. A., & Ebadi, M. J. (2023). A truthful and budget-balanced double auction model for resource allocation in cloud computing. *Soft Computing*. https://doi.org/10.1007/s00500-023-08081-4

Alipour, P., & Bastani, A. F. (2023). Value-at-Risk-Based Portfolio Insurance: Performance Evaluation and Benchmarking Against CPPI in a Markov-Modulated Regime-Switching Market. arXiv preprint arXiv:2305.12539

Arora, N., & Kaur, P. D. (2020). A Bolasso based consistent feature selection enabled random forest classification algorithm: An application to credit risk assessment. *Applied Soft Computing, 86*, 105936.

Bachute, M. R., & Subhedar, J. M. (2021). Autonomous driving architectures: Insights of machine learning and deep learning algorithms. *Machine Learning with Applications, 6*, 1172.

Bhatia, S., Sharma, P., Burman, R., Hazari, S., & Hande, R. (2017). Credit scoring using machine learning techniques. *International Journal of Computers and Applications, 161*, 1–4.

Chen, Z., Zhu, W., Feng, H., & Luo, H. (2022). Changes in corporate social responsibility efficiency in Chinese food industry brought by COVID-19 pandemic—a study with the super-efficiency DEA-Malmquist-Tobit model. *Frontiers in Public Health, 10*, 875030.

Dehghani, F., & Larijani, A. (2023). An Algorithm for Predicting Stock Market's Index Based on MID Algorithm and Neural Network. Available at SSRN 4448033.

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of online earning and an application to boosting. *Journal of Computer and System Sciences, 55*, 119–139.

Gao, H., Liu, Z., & Yang, C. C. (2023). Individual investors' trading behavior and gender difference in tolerance of sex crimes: Evidence from a natural experiment. *Journal of Empirical Finance, 73*, 349–368.

Guo, H., Zhou, J., & Koopialipoor, M. (2019). Deep neural network and whale optimization algorithm to assess flyrock induced by blasting. *Engineering with Computers* 1–14.

Harris, T. (2015). Credit scoring using the clustered support vector machine. *Expert Systems with Applications, 42*, 741–750.

He, C., Huang, K., Lin, J., Wang, T., & Zhang, Z. (2023). Explain systemic risk of commodity futures market by dynamic network. *International Review of Financial Analysis, 88*, 102658.

Jiang, Z., & Xu, C. (2023). Disrupting the technology innovation efficiency of manufacturing enterprises through digital technology promotion: An evidence of 5G technology construction in China. *IEEE Transactions on Engineering Management*. https://doi.org/10.1109/TEM.2023.3261940

Kang, Y., Chen, L., Jia, N., Wei, W., Deng, J., & Qian, H. (2022). A CWGAN-GP-based multi-task learning model for consumer credit scoring. *Expert Systems with Applications, 206*, 117650.

Kuppili, V., Tripathi, D., & Reddy Edla, D. R. (2020). Credit score classification using spiking extreme learning machine. *Computational Intelligence, 36*, 402–426.

Li, X., & Sun, Y. (2020). Stock intelligent investment strategy based on support vector machine parameter optimization algorithm. *Neural Computing and Applications, 32*(6), 1765–1775.

Li, X., & Sun, Y. (2021). Application of RBF neural network optimal segmentation algorithm in credit rating. *Neural Computing and Applications, 33*(14), 8227–8235.

Li, G., Ma, H. D., Liu, R. Y., Shen, M. D., & Zhang, K. X. (2021). A two-stage hybrid default discriminant model based on Deep Forest. *Entropy, 23*, 582.

Li, Z., Zhou, X., & Huang, S. (2021). Managing skill certification in online outsourcing platforms: A perspective of buyer-determined reverse auctions. *International Journal of Production Economics, 238*, 108166.

Liu, W., Fan, H., & Xia, M. (2022). Multi-grained and multi-layered gradient boosting decision tree for credit scoring. *Applied Intelligence, 52*, 5325–5341.

Liu, X., Zhou, G., Kong, M., Yin, Z., Li, X., Yin, L., & Zheng, W. (2023). Developing multi-labelled corpus of twitter short texts: A semi-automatic method. *Systems, 11*(8), 390.

Liu, X., Wang, S., Lu, S., Yin, Z., Li, X., Yin, L., Tian, J., & Zheng, W. (2023a). Adapting feature selection algorithms for the classification of Chinese texts. *Systems, 11*(9), 483.

Liu, X., Shi, T., Zhou, G., Liu, M., Yin, Z., Yin, L., & Zheng, W. (2023b). Emotion classification for short texts: An improved multi-label method. *Humanities and Social Sciences Communications, 10*(1), 306.

Luo, J., Wang, Y., & Li, G. (2023). The innovation effect of administrative hierarchy on intercity connection: The machine learning of twin cities. *Journal of Innovation & Knowledge, 8*(1), 100293.

Mandala, I. G. N. N., Nawangpalupi, C. B., & Praktikto, F. R. (2012). Assessing credit risk: An application of data mining in a rural bank. *Procedia Econ. Financ., 4*, 406–412.

Martens, D., Baesens, B., Van Gestel, T., & Vanthienen, J. (2007). Comprehensible credit scoring models using support vector machines. *European Journal of Operational Research, 183*, 1488–1497.

Munkhdalai, L., Ryu, K., Namsrai, O.-E., & Theera-Umpon, N. (2021). A partially interpretable adaptive softmax regression for credit scoring. *Applied Sciences, 11*, 3227.
Nikulin, E. E., & Pekhterev, A. A. (2021). Turbulence on financial markets and a model of a multiplicative cascade of volatility. *Mathematical Models and Computer Simulations, 13*(4), 660–666. https://doi.org/10.1134/S2070048221040177
Olivier, N., & Lieven, D. M. (2019). Debt rating downgrades of financial institutions: Causality tests on singleissue CDS and iTraxx. *Quantitative Finance, 19*(12), 1975–1993. https://doi.org/10.1080/14697688.2019.1619933
Ouyang, Z. S., Yang, X. T., & Lai, Y. Z. (2021). Systemic financial risk early warning of financial market in China using Attention-LSTM model. *The North American Journal of Economics and Finance, 56*, 1917. https://doi.org/10.1016/j.najef.2021.101383
Peng, Y., Zhao, Y., & Hu, J. (2023). On the role of community structure in evolution of opinion formation: A new bounded confidence opinion dynamics. *Information Sciences, 621*, 672–690.
Qian, X. Y., & Olsen, T. L. (2020). Financial and risk management in agricultural cooperatives with application to the milk industry in New Zealand. *International Journal of Production Research, 62*(1–31), 2020.
Radović, O., Marinković, S., & Radojičić, J. (2021). Credit scoring with an ensemble deep learning classification methods—Comparison with traditional methods. *Facta Universitatis, Series: Economics and Organization, 18*, 29–43.
Razmi, S. F., Torki, L., Razmi, M. J., & Dowlatabadi, E. M. (2022). The indirect effects of oil price on consumption through assets. *International Journal of Energy Economics and Policy, 12*(1), 236–242.
Rostami, M. R., Alipour, P., & Behzadi, A. (1999). Analyzing the causal relations between trading volume and stock returns and between trading volume and return volatility in Tehran Stock Exchange. *Iranian Journal of Finance, 2*(4), 27–40.
Sha, Z., Hu, L., & Rouyendegh, B. D. (2020). Deep learning and optimization algorithms for automatic breast cancer detection. *International Journal of Imaging Systems and Technology, 30*(2), 495–506.
Sinanc, D., Demirezen, U., & Sağıroğlu, Ş. (2021). Explainable credit card fraud detection with image conversion. *ADCAIJ, 2021*(10), 63–76.
Soares, C., Figueiredo, A., & Figueiredo, F. (2021). The subprime crisis and its influence on public, private and financial sectors in European countries. *Journal of East-West Business, 27*(1), 92–108. https://doi.org/10.1080/10669868.2020.1830913
Tehranian, K. (2023). Can Machine Learning Catch Economic Recessions Using Economic and Market Sentiments? arXiv preprint arXiv:2308.16200
Tripathi, D., Edla, D. R., Kuppili, V., & Bablani, A. (2020). Evolutionary extreme learning machine with novel activation function for credit scoring. *Engineering Applications of Artificial Intelligence, 96*, 103980.
Tripathi, D., Edla, D. R., Bablani, A., Shukla, A. K., & Reddy, B. R. (2021). Experimental analysis of machine learning methods for credit score classification. *Progress in Artificial Intelligence, 10*, 217–243.
Trivedi, S. K. (2020). A study on credit scoring modeling with different feature selection and machine learning approaches. *Technology in Society, 63*, 101413.
Wang, D., Zhang, Z., Bai, R., & Mao, Y. (2018). A hybrid system with filter approach and multiple population genetic algorithm for feature selection in credit scoring. *Journal of Computational and Applied Mathematics, 329*, 307–321.
Wang, H., Wu, Z., Chen, J., & Chen, L. (2022). Evaluation of road traffic noise exposure considering differential crowd characteristics. *Transportation Research Part D: Transport and Environment, 105*, 103250.
Wang, K., Hu, Y., Zhou, J., & Hu, F. (2023). Fintech, financial constraints and OFDI: Evidence from China. *Global Economic Review, 52*(4), 326–345.
Xu, A., Qiu, K., & Zhu, Y. (2023). The measurements and decomposition of innovation inequality: Based on Industry − University − Research perspective. *Journal of Business Research, 157*, 113556.
Xuemin, Z., Ying, R., Zenggang, X., Haitao, D., Fang, X., & Yuan, L. (2023). Resource-constrained and socially selfish-based incentive algorithm for socially aware networks. *Journal of Signal Processing Systems, 95*(12), 1439–1453.
Yin, Y., Zhang, R., & Su, Q. (2023). Threat assessment of aerial targets based on improved GRA-TOPSIS method and three-way decisions. *Mathematical Biosciences and Engineering, 20*(7), 13250–13266.
Zhang, T., & Chi, G. (2020). A heterogeneous ensemble credit scoring model based on adaptive classifier selection: An application on imbalanced data. *International Journal of Finance & Economics, 26*, 4372–4385.
Zhang, W., Yang, D., Zhang, S., Ablanedo-Rosas, J. H., Wu, X., & Lou, Y. (2021). A novel multi-stage ensemble model with enhanced outlier adaptation for credit scoring. *Expert Systems with Applications, 165*, 113872.