2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

# Deep Learning-Based Credit Score Prediction: Hybrid LSTM-GRU Model

Golnaz Sababipour ASL, Kiarash Shamsi, Ruppa K. Thulasiram, Cuneyt Akcora, Carson Leung
*Department of Computer Science*
*University of Manitoba*
Winnipeg, Canada
sababipg@myumanitoba.ca, shamsik1@myumanitoba.ca, tulsi.thulasiram@umanitoba.ca,
cuneyt.akcora@umanitoba.ca, kleung@cs.umanitoba.ca

***Abstract*—Credit score prediction is a crucial task in financial industry, as it helps lenders and financial institutions evaluate the creditworthiness of borrowers and manage credit risk. In this work, we present a comparative study of deep learning (DL)-based credit score prediction models. To achieve this objective, we compare the performance of DL models against traditional methods in credit scoring. We train and test the models using a real-world dataset of credit histories, containing various features such as credit card balances, payment history, and employment status. Our experimental results show that the hybrid LSTM-GRU model outperform both the LSTM and GRU models in credit score prediction, as well as traditional methods. The hybrid LSTM-GRU model demonstrates higher accuracy and better predictive power, indicating its potential for improving credit scoring models in the financial industry.**

***Index Terms*—Financial Intelligence, Credit Score, Deep Learning, RNNs, Hybrid model**

## I. Introduction

Approximately 60% of the risk faced by banks is associated with credit risks [1]. Credit risk is defined as the likelihood of a financial loss stemming from a borrower's inability to repay a loan. This type of risk relates to the possibility of a lender not receiving the principal and interest owed, resulting in a disruption of cash flows and increased costs for debt collection. For lenders and investors who assume credit risk, interest payments from borrowers or issuers of debt obligations serve as form of compensation. To mitigate credit risk, lenders can analyze factors about a borrower's creditworthiness. While it is impossible to predict precisely which borrowers will default on their loans, assessing and managing credit risk can reduce the severity of potential losses. The widespread adoption of financial services has drawn the interest of scholars to credit risk management. As a result, researchers have developed models with the dual goals of mitigating financial risks and maximizing associated profits. The predictions generated by credit scoring models have become a crucial aspect of financial institutions. Credit scoring is considered a traditional decision-making model that seeks to evaluate the risk related to credit products, including credit cards and loans. Historical data from an applicant is used in this evaluation, and the credit score produced can aid lenders in making well-informed decisions about extending credit. To decrease credit risk, financial institutions are increasingly utilizing various risk assessment tools and techniques, and they often use statistical analysis of customer credit data to identify potential defaulters [2]. Credit score has various use cases including lending decisions, interest rates determination, insurance rate arbitration, employment screening, rental applications, and utilities and services. These use cases highlight the importance of credit score prediction as a tool for evaluating an individual's creditworthiness and assessing their risk of default. By using credit score prediction models, financial institutions can make informed decisions about extending credit and managing credit risk [3], which can ultimately impact their profitability and financial stability. The purpose of this study is to enhance the prediction of credit scores for current customers of financial institutions, thereby assisting in the effective allocation of resources to non-defaulting individuals.

DL models are increasingly being utilized for credit score prediction [4]–[7]. The models become significant in their capability to capture complex, non-linear, and sequential relationships between variables, automatically extract features from data, handle larger and more complex datasets, and provide flexibility for customization to specific lending environments [8]. Utilizing DL algorithms to forecast credit scores has the potential to enhance overall accuracy and efficacy.

In the field of credit score prediction, recurrent neural network (RNN) models have gained significant attention. RNNs are a category of neural networks developed specifically for handling sequential data, making them particularly well-suited for analyzing credit-related information that exhibits temporal dependencies, such as transaction histories or time series data. Unlike traditional statistical models, RNNs have the ability to capture and retain information from previous time steps, allowing them to uncover complex patterns and long-term dependencies that may exist in credit data over time [9]. By utilizing recurrent connections, RNNs can retain information about past credit behavior and use it to make informed predictions about future creditworthiness. This is crucial for assessing an individual's credit risk accurately [10]–[12]. Given the increasing demand for precise credit score assessment and the availability of vast amounts of credit-related data, the use of RNN models and it's hybrid extensions [13] in credit score prediction has gained prominence.

978-0-7381-4408-5/23/$31.00 ©2023 IEEE

In this study we implemented several DL models including long short-term memory (LSTM), convolutional neural networks (CNNs), gated recurrent unit (GRU), and a CNN-LSTM hybrid model that have been shown to improve the performance [14]. We have also proposed an LSTM-GRU hybrid model to evaluate the performance of existing models for credit prediction.

## II. Related Work

In general, the study of credit scores can be categorized into two primary domains: the first domain centers around the advancement of accurate credit score computation techniques and the exploration of strategies for enhancing them, whereas the second domain centers around the comparison and assessment of various credit score models in order to gauge their efficacy.

### A. Credit score prediction

Ala'raj et al. [15] discussed the importance of models to capture credit score in assessing and reducing bank losses, and proposed a bidirectional LSTM model for predicting missed credit card payments by customers. Their aim was to forecast the behavior of credit card customers regarding the probability of missing payments, both for individual and consecutive cases. The authors demonstrated that the LSTM model outperforms the traditional models in terms of accuracy, area under curve(AUC), Kolmogorov–Smirnov test, H-measure, calibration curves, Brier score, and the McNemar test. Adisa et al. [16] explained the utilization of LSTM for credit scoring prediction. An optimization approach was employed, employing a genetic algorithm (GA) to identify the most favorable parameters for the LSTM model. The findings demonstrated that the optimized LSTM model surpassed both individual classifiers and ensemble models in terms of accuracy and loss when it came to predictions. The study concluded that the hybrid LSTM model performs better than all other models employed in the research. Kumar et al. [5] proposed an approach to predict credit scoring for customers in the financial industry, utilizing a combination of DL and k-Means algorithm. The proposed scheme incorporated feature selection, DL, and decision tree classification in order to effectively predict credit scores. The study's results demonstrate that the proposed model performs well in predicting credit scores for existing customers and can aid lenders in allocating funds in the finance industry. Dastile and Celik [17] proposed a novel DL model for credit scoring. Their method converted tabular datasets into images, where each pixel represented a feature bin from the tabular data. The authors used state-of-the-art explanation methods to provide insights into the predictions made by the 2D CNNs, and showed that the trained CNN performed better when compared to results in the published literature.

### B. Comparing the scoring models

Sakri [18] compared the classification power of Deep Neural Networks (DNN) and Gradient Boosting Machines (GBM) for credit risk prediction. Three credit risk datasets were used to train and evaluate GBM, DNN+ReLU, DNN+Maxout, and DNN+Tanh classifiers. The accuracy of the models was computed using the area under the receiver operating characteristics curve (AUC ROC). The results showed that GBM was faster and more potent than DNN because of its lower processing requirements and higher accuracy. Trivedi [19] used publicly available German credit data and compared feature selection techniques and five ML classifiers. The Random Forest classifier with the Chi-Square feature selection method was found to be the best combination, reducing false positives and false negatives. Decision tree was also found to be a good performer. Zhu et al. [20] introduced a novel model that combined CNN with the Relief algorithm. The model's performance was compared with a hybrid approach involving logistic regression and random forest, using a real-world Chinese consumer finance company dataset. The results showed that the Relief-CNN hybrid model outperformed the benchmark algorithms. The researchers expressed their strong conviction in the capability of deep learning techniques to offer robust assistance in credit scoring. Golbayani et al. [21] analyzed the performance of four different NN architectures including MLP, CNN, CNN2D, and LSTM in the energy, financial, and healthcare sectors in the United States. The study aimed to improve the application of ML algorithms in credit assessment and addressed the main questions. The results were analyzed using ANOVA and multiple comparison testing procedures.

The current study includes both domains. We first explore various models and their performance. Next, we propose a new hybrid model, which combines LSTM and GRU.

## III. Methodology

### A. Data

The loan dataset used for this study contains comprehensive information on consumer loans issued by the Lending Club in the US, spanning from 2007 to 2014 [22]. The dataset comprises 75 features, including current loan status and various borrower-related attributes such as employment length, credit score, and debt-to-income ratio. The main use case of this dataset is to predict the likelihood of a borrower defaulting on their loan. Figure 1 shows the distribution of the loan_status labels in the dataset. The majority of loans in the dataset are in the Fully Paid with label 1, while a minority are in the Charged Off with label 0. The final decision regarding lending money is based on this informed label.

We performed five preprocessing steps to prepare the loan dataset for the learning process, as described below:

- **Handling Missing Values:** We dropped columns that had more than 80% null values. These columns were deemed to be of little use in the analysis and could potentially mislead our models. Description of the datasets used in the study after applying first step demonstrates in Table I. The number of features after the data processing step is 42, comprising 34 categorical features and 8 numerical features.

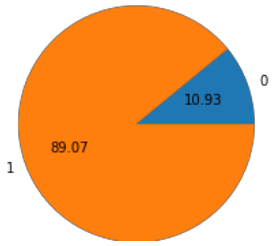


Fig. 1. The distribution of loan_status labels in the dataset.

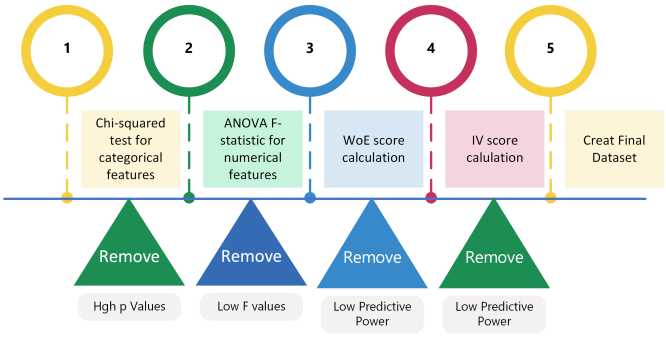


Fig. 2. Steps of feature selection process.

- **Reshape:** We performed feature reshaping and format/type conversion on the loan dataset. Some features, such as interest rates and loan amounts, were in string format and needed conversion to numerical format, while some categorical strings like employment length were encoded into numerical values.
- **Feature Selection:** To identify relevant and informative features, we did a feature selection for our modeling tasks. We employed several techniques, including the chi-squared test, analysis of variance (ANOVA) F-statistic correlation, and WoE and IV (Weight of Evidence and Information Value) [23]. The result of chi-squared test demonstrated the features were significantly associated with the target variable. We selected the first four features based on their p-values, which were almost 0. These features included grade, home ownership, verification_status, and purpose. We conducted an Anova on 34 features, and the results revealed that a significant number of the listed numerical features play a crucial role in predicting or explaining the target variable. This was evidenced by their notably low p-values and high F-scores. Consequently, we selected the top 20 features. The WoE and IV technique is particularly useful for measuring the predictive power of independent variables in relation to the target variable [24], while the chi-squared test and ANOVA F-statistic correlation helped us to identify the features that had the most impact on our target variable [25]. Using these techniques, we were able to increase the accuracy of our models slightly. We employed WoE and IV analysis on all numerical and categorical features. A higher IV value signifies a greater predictive power of the feature regarding the target variable. Features with IV values less than 0.003 were removed from consideration, as they did not contribute significantly to the predictive capability. Figure 2 shows the feature selection steps.
- **Categorical to Numerical:** After feature selection, we converted the remaining categorical features in the loan dataset to numerical format using one-hot encoding.
- **Data Split:** After preprocessing the dataset, we splitted it into 80% training data and 20% testing data to evaluate the performance of our ML models.

TABLE I
DESCRIPTION OF THE DATASETS

| Train Size | Test Size | Number of features | |
|---|---|---|---|
| | | Before processing | After processing |
| 4373028 | 93257 | 75 | 42 |

### B. ML and DL Baseline Models

We aim to create baseline evaluations to provide a clear starting point for measuring the effectiveness of the ML model that we are developing. Evaluating our models against established methods allows us to assess their performance and identify potential areas for optimization. We use a combination of ML and DL models as baselines to compare their performance in predicting a specific target variable. We implement Random Forest, SVM, and XGBoost as our ML models, which are well-established algorithms in the field [26]–[28]. Then, we explore the capabilities of DL models, including CNN and RNN based models. For RNN, we used both GRU and LSTM and compare their performance.

Using both ML and DL models, we aim to show the strengths and limitations of each method. ML models are generally faster to train and easier to interpret, while DL models can handle complex relationships between inputs and outputs. Furthermore, we create a hybrid model by combining the best-performing DL models. This approach can potentially improve the overall performance of the model and provide more accurate predictions.

### C. Model Creation

The purpose of this study is to build a hybrid model by combining the strengths of two different well studied models of neural networks, LSTM and GRU. Our goal was to improve the accuracy of binary classification tasks by using the advantages of each network.

The hybrid model has three layers. The first layer is an LSTM layer. This layer has 64 units and return sequences set to true. The output of this layer is fed into the second layer, which is a GRU layer with 32 units and an activation function of ReLU. The third and final layer is a dense layer with a single output unit and employs a sigmoid activation function. The loss function used in this study is binary cross-entropy loss and model optimization is done using Adam optimizer. The ROC AUC, F1 score, precision, and recall metrics are employed for assessing our model's performance.

TABLE II
CONFUSION MATRIX FOR THE RANDOM FOREST MODEL.

| | | Actual Value | |
|---|---|---|---|
| | | Positive | Negative |
| Predictions | Positive | 9223 | 971 |
| | Negative | 193 | 82870 |

TABLE III
CONFUSION MATRIX FOR THE LSTM MODEL.

| | | Actual Value | |
|---|---|---|---|
| | | Positive | Negative |
| Predictions | Positive | 6992 | 3202 |
| | Negative | 672 | 82391 |

To optimize our model, we use grid search to tune the hyperparameters. We experiment with different combinations of hyperparameters, including the number of units in the LSTM and GRU layers, learning rate, epoch and batch size. Our goal is to identify the optimal hyperparameters that would result in the highest ROC AUC score. By fine-tuning the hyperparameters using grid search, we are able to identify the optimal configuration of our hybrid model, which achieved superior performance compared to the individual LSTM and GRU models.

## IV. EXPERIMENTAL RESULTS

Our experimental result is presented on machine-learning models for a specific task. We evaluate the performance of our models using four widely used metrics: AUC, F1 score, precision, and recall. AUC is a measure of the overall performance of a model, which indicates how well the model can distinguish between positive and negative samples. F1 score is the harmonic mean of recall and precision, which is usual metric for (binary) classification problems. Recall measures the part of true positive results among the actual positive results while precision measures the part of true positive results among all predicted positive results. These metrics are important in evaluating the effectiveness and accuracy of ML models for various tasks.

To improve our understanding of our models' performance, we created confusion matrices for Random Forest, LSTM, GRU, and LSTM-GRU, as demonstrated in Table II, Table III, Table IV, and Table V, respectively. These tables provide a detailed breakdown the performance of the model in predicting true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). By examining these results, we can see that our hybrid LSTM-GRU model outperformed the individual LSTM model and can improve the GRU model slightly in terms of accuracy, precision, and recall. These results suggest that our hybrid model can effectively leverage the strengths of both LSTM and GRU architectures to achieve better predictive performance.

TABLE IV
CONFUSION MATRIX FOR THE GRU MODEL.

| | | Actual Value | |
|---|---|---|---|
| | | Positive | Negative |
| Predictions | Positive | 9678 | 516 |
| | Negative | 85 | 82978 |

TABLE V
CONFUSION MATRIX FOR THE LSTM-GRU MODEL.

| | | Actual Value | |
|---|---|---|---|
| | | Positive | Negative |
| Predictions | Positive | 9685 | 509 |
| | Negative | 70 | 82993 |

Table VI presents the performance metrics of different ML models, including Random Forest, SVM, XGBoost, CNN, GRU, LSTM, CNN-LSTM, and LSTM-GRU, in predicting the target variable. We assessed the performance of each model by considering AUC, F1 score, precision, and recall. The results indicate that the LSTM-GRU model outperformed other models in all performance metrics, except for precision. LSTM-GRU achieves an AUC score of 0.98, F1 score of 0.98, recall score of 0.99, and precision score of 0.96. AUC is considered the most important metric in credit score prediction, as it offers an assessment of the model's capacity to differentiate between good and bad credit applicants. The LSTM-GRU model achieves the highest AUC score of 0.98, indicating its superior performance in predicting credit scores compared to other models.

It is worth noting that while the CNN model has a relatively high precision score of 0.97, it suffers from low recall and F1 scores, suggesting that it fails to identify a significant number of positive cases. Additionally, the CNN-LSTM model achieves the lowest performance scores among all models, which may be attributed to the difficulty of training hybrid models that combine two different architectures.

Overall, the results demonstrate the potential of DL models, especially LSTM-GRU, in credit score prediction, offering a more accurate and powerful approach compared to traditional methods.

### A. Explaining the models

Explainable Artificial Intelligence (XAI) is a crucial area of research in DL that aims to provide transparency and interpretability in models' decision-making processes [29]. XAI techniques such as feature contributions can help identify the most critical features that contribute to a model's decision-making process, thus improving the model's accuracy and preventing biases [30].

Figure 3 depicts the importance scores of our dataset. The results show that a few features have considerably high importance, whereas most other features are negligible in their contribution to the model.

## V. CONCLUSION

Credit score prediction is a complex problem that requires the analysis of a vast amount of data, including the borrower's financial history, credit behavior, and other factors that may impact their creditworthiness. In recent years, DL

TABLE VI
PERFORMANCE METRICS FOR DIFFERENT MACHINE LEARNING MODELS

| Metrics | Random Forest | SVM | XGBOOST | CNN | GRU | LSTM | CNN-LSTM | LSTM-GRU |
|---|---|---|---|---|---|---|---|---|
| AUC | 0.95 | 0.60 | 0.96 | 0.95 | 0.97 | 0.83 | 0.50 | **0.98** |
| F1 score | 0.96 | 0.65 | 0.98 | 0.96 | 0.98 | 0.88 | 0.47 | 0.98 |
| Precision | 0.98 | 0.91 | 0.99 | 0.97 | 0.99 | 0.96 | 0.89 | 0.99 |
| Recall | 0.99 | 0.99 | 0.99 | 0.97 | 0.99 | 0.99 | 0.99 | 0.99 |

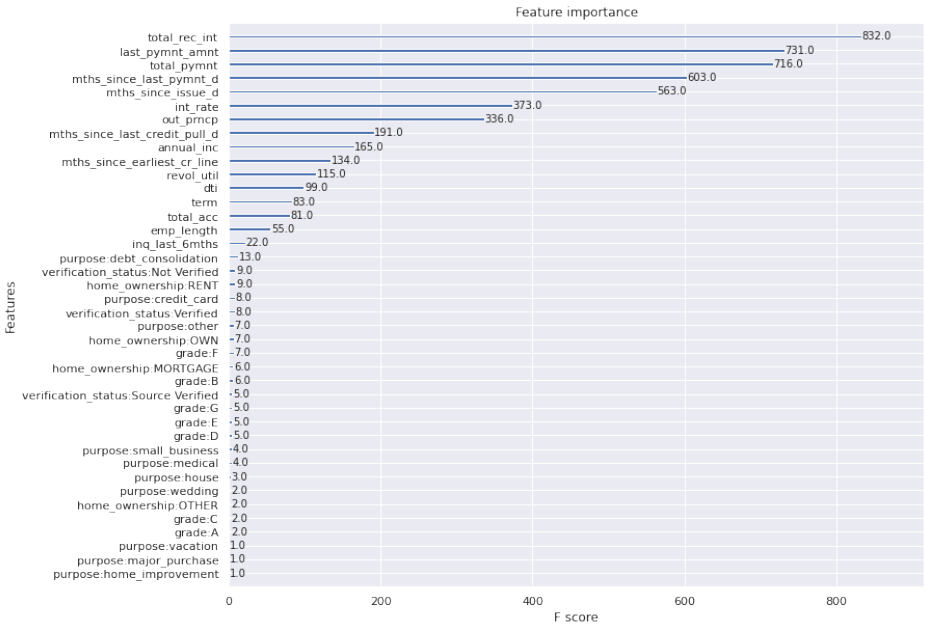


Fig. 3. Feature importance of the dataset.

techniques have shown promise in credit score prediction, offering improved accuracy and predictive power over traditional methods. The principle aim of our study is to determine which DL model performs best in predicting credit scores. The use of hybrid LSTM-GRU models has shown significant improvements in the accuracy and predictive power of credit score prediction compared to traditional methods. The experimental results indicate that the LSTM-GRU hybrid model significantly outperformed the LSTM model and shows a slight improvement over the GRU model. These findings emphasize the importance of carefully selecting the appropriate DL model for credit scoring, as the performance and predictive power of different models can vary significantly. This study has contributed to the growing body of research on DL-based credit score prediction and highlights the potential of these techniques in improving financial decision-making. As the availability of data and computing power continues to increase, it is likely that DL will become more and more crucial in credit scoring and other financial tasks. Overall, the superior performance of the LSTM-GRU model in capturing complex patterns and relationships in credit data suggested its potential for practical applications in credit scoring.

## ACKNOWLEDGMENT

The first author acknowledges the Research Assistantship from Prof. Thulasiram and Graduate Enhancement of Tri-agency Stipends (GETS) from University of Manitoba. The second author acknowledges the Research Assistantship from Prof. Akcora and Graduate Enhancement of Tri-agency Stipends (GETS) and the Faculty of Graduate Studies (FGS) from University of Manitoba. The last three authors acknowledge the Natural Sciences and Engineering Research Council (NSERC) of Canada for Discovery Grants.

## REFERENCES

[1] K. Buehler, A. Freeman, and R. Hulme, "The new arsenal of risk management," *Harvard business review*, vol. 86, pp. 92–100+137, 09 2008.

[2] N. Darapaneni, A. Kumar, A. Dixet, M. Suriyanarayanan, S. Srivastava, and A. R. Paduri, "Loan prediction software for financial institutions," in *2022 Interdisciplinary Research in Technology and Management (IRTM)*, pp. 1–8, IEEE, 2022.

[3] X.-L. Li and Y. Zhong, "An overview of personal credit scoring: techniques and future work," 2012.

[4] V.-S. Ha and H.-N. Nguyen, "Credit scoring with a feature selection approach based deep learning," in *MATEC web of conferences*, vol. 54, p. 05004, EDP Sciences, 2016.

[5] A. Kumar, D. Shanthi, and P. Bhattacharya, "Credit score prediction system using deep learning and k-means algorithms," in *Journal of Physics: Conference Series*, vol. 1998, p. 012027, IOP Publishing, 2021.

[6] C. Wang, D. Han, Q. Liu, and S. Luo, "A deep learning approach for credit scoring of peer-to-peer lending using attention mechanism lstm," *IEEE Access*, vol. 7, pp. 2161–2168, 2018.

[7] S. Pol, S. S. Ambekar, *et al.*, "Predicting credit ratings using deep learning models–an analysis of the indian it industry," *Australasian Accounting, Business and Finance Journal*, vol. 16, no. 5, pp. 38–51, 2022.

[8] P. Golbayani, I. Florescu, and R. Chatterjee, "A comparative study of forecasting corporate credit ratings using neural networks, support vector machines, and decision trees," *The North American Journal of Economics and Finance*, vol. 54, p. 101251, 2020.

[9] P. Du and H. Shu, "Exploration of financial market credit scoring and risk management and prediction using deep learning and bionic algorithm," *Journal of Global Information Management (JGIM)*, vol. 30, no. 9, pp. 1–29, 2022.

[10] D. Babaev, M. Savchenko, A. Tuzhilin, and D. Umerenkov, "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, 2019.

[11] T.-C. Hsu, S.-T. Liou, Y.-P. Wang, Y.-S. Huang, *et al.*, "Enhanced recurrent neural network for combining static and dynamic features for credit card default prediction," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1572–1576, IEEE, 2019.

[12] M. Ala'raj, M. F. Abbod, M. Majdalawieh, and L. Jum'a, "A deep learning model for behavioural credit scoring in banks," *Neural Computing and Applications*, pp. 1–28, 2022.

[13] J. Li, C. Xu, B. Feng, and H. Zhao, "Credit risk prediction model for listed companies based on cnn-lstm and attention mechanism," *Electronics*, vol. 12, no. 7, p. 1643, 2023.

[14] M. A. Hossain, R. Karim, R. Thulasiram, N. D. Bruce, and Y. Wang, "Hybrid deep learning model for stock price prediction," in *2018 ieee symposium series on computational intelligence (ssci)*, pp. 1837–1844, IEEE, 2018.

[15] M. Ala'raj, M. F. Abbod, and M. Majdalawieh, "Modelling customers credit card behaviour using bidirectional lstm neural networks," *Journal of Big Data*, vol. 8, no. 1, pp. 1–27, 2021.

[16] J. Adisa, S. Ojo, P. Owolawi, A. Pretorius, and S. O. Ojo, "Credit score prediction using genetic algorithm-lstm technique," in *2022 Conference on Information Communications Technology and Society (ICTAS)*, pp. 1–6, IEEE, 2022.

[17] X. Dastile and T. Celik, "Making deep learning-based predictions for credit scoring explainable," *IEEE Access*, vol. 9, pp. 50426–50440, 2021.

[18] S. Sakri, "Assessment of deep neural network and gradient boosting machines for credit risk prediction accuracy," in *2022 14th International Conference on Computational Intelligence and Communication Networks (CICN)*, pp. 1–7, IEEE, 2022.

[19] S. K. Trivedi, "A study on credit scoring modeling with different feature selection and machine learning approaches," *Technology in Society*, vol. 63, p. 101413, 2020.

[20] B. Zhu, W. Yang, H. Wang, and Y. Yuan, "A hybrid deep learning model for consumer credit scoring," in *2018 international conference on artificial intelligence and big data (ICAIBD)*, pp. 205–208, IEEE, 2018.

[21] M. Gerndt, *Automatic Parallelization for Distributed-Memory Multiprocessing Systems*. PhD thesis, University of Bonn, Bonn, Germany, Dec. 1989.

[22] L. C. site, "Lending club dataset 2007-2015." Kaggle: A platform for data science competitions, 2015.

[23] B. Baesens, *Analytics in a big data world: The essential guide to data science and its applications*. John Wiley & Sons, 2014.

[24] S. Ruiz, P. Gomes, L. Rodrigues, and J. Gama, "Assembled feature selection for credit scoring in microfinance with non-traditional features," in *Discovery Science: 23rd International Conference, DS 2020, Thessaloniki, Greece, October 19–21, 2020, Proceedings 23*, pp. 207–216, Springer, 2020.

[25] A. Siham, S. Sara, and A. Abdellah, "Feature selection based on machine learning for credit scoring: An evaluation of filter and embedded methods," in *2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, pp. 1–6, IEEE, 2021.

[26] C. Chen, S. Yokoyama, T. Yamashita, and H. Kawamura, "Application of xgboost to credit scoring," *Information processing Society Of Japan*, vol. 194, no. 11, pp. 1–8, 2019.

[27] T. Bellotti and J. Crook, "Support vector machines for credit scoring and discovery of significant features," *Expert systems with applications*, vol. 36, no. 2, pp. 3302–3308, 2009.

[28] L. Zhu, D. Qiu, D. Ergu, C. Ying, and K. Liu, "A study on predicting loan default based on the random forest algorithm," *Procedia Computer Science*, vol. 162, pp. 503–513, 2019.

[29] S. Ali, T. Abuhmed, S. El-Sappagh, K. Muhammad, J. M. Alonso-Moral, R. Confalonieri, R. Guidotti, J. Del Ser, N. Díaz-Rodríguez, and F. Herrera, "Explainable artificial intelligence (xai): What we know and what is left to attain trustworthy artificial intelligence," *Information Fusion*, p. 101805, 2023.

[30] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G.-Z. Yang, "Xai—explainable artificial intelligence," *Science robotics*, vol. 4, no. 37, p. eaay7120, 2019.