2023 XXIX International Conference on Information, Communication and Automation Technologies (ICAT) | 979-8-3503-9983-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAT57854.2023.10171259

Proposal of a model for credit risk prediction based on deep learning methods and SMOTE techniques for imbalanced dataset

Adaleta Gicić
Faculty of Electrical Engineering
University of Sarajevo
Sarajevo,Bosnia and Herzegovina
adaleta.gicic@etf.unsa.ba

Dženana Đonko
Faculty of Electrical Engineering
University of Sarajevo
Sarajevo,Bosnia and Herzegovina
ddonko@etf.unsa.ba

***Abstract*— Implementation of credit scoring models is a demanding task and crucial for risk management. Wrong decisions can significantly affect revenue, increase costs, and can lead to bankruptcy. Together with the improvement of machine learning algorithms over time, credit models based on novel algorithms have also improved and evolved. In this work, novel deep neural architectures, Stacked LSTM, and Stacked BiLSTM combined with SMOTE oversampling technique for the imbalanced dataset were developed and analyzed. The reason for the lack of publications that utilize Stacked LSTM-based models in credit scoring lies exactly in the fact that the deep learning algorithm is tailored to predict the next value of the time series, and credit scoring is a classification problem. The challenge and novelty of this approach involved the necessary adaptation of the credit scoring dataset to suit the time sequence nature of LSTM-based models. This was particularly crucial as, in practical credit scoring datasets, instances are not correlated nor time dependent. Moreover, the application of SMOTE to the newly constructed three-dimensional array served as an additional refinement step. The results show that techniques and novel approaches used in this study improved the performance of credit score prediction.**

Keywords*— *Deep Learning, Credit Scoring, Stacked LSTM, Stacked BiLSTM, SMOTE, Credit Risk Management

I. INTRODUCTION

Credit scoring is an important tool in credit risk management. It is a method of assigning the sum of points to the borrower that represents a numerical value that shows how likely the borrower is to be late in repaying the loan. It decides if an applicant belongs to a good or bad class, although three class models can also be found in practice and publications [1]. In its essence, credit scoring is a classification problem [2]. Credit scoring techniques can be divided into statistical and machine learning techniques. Classical credit scoring models such as statistical techniques are limited as they often require restrictive statistical assumptions that are seldom satisfied in real cases. Machine learning modeling on the contrary does not need domain experts to develop rules. Algorithms are developed and applied to discover and recognize patterns in data to use them in predictive modeling.

In recent years we witness limitless opportunities mostly due to expansion in the artificial intelligence field. Manual credit approval decision assessments were time-consuming and prone to errors and therefore were mostly replaced by machine learning-based prediction systems. Deep learning has received much attention in recent years, but it has not been implemented so intensively in credit scoring compared to other financial domains [3].

This paper will fill this gap by applying Stacked LSTM (Long Short-term Memory Networks) and BiLSTM (Bidirectional LSTM) deep learning algorithm boosted with SMOTE (Synthetic Minority Oversampling Technique) on a well-known German credit dataset.

The novelty of this study lies in the successful adaptation of credit scoring data to effectively utilize deep neural architectures, specifically those that operate on a 3D array structure representing time sequence data. However, it should be noted that the non-dependency of rows within the array is a critical consideration, as the inherent nature of the data does not establish any interdependence among them. The central concept revolves around a unique and innovative modeling approach that successfully achieves the exclusion of previous events from being incorporated during the model's training process. In this methodology, emphasis is put solely on the attributes pertaining to the observed customer, treating them as the only sequence of events to be considered.

When employing matrices to structure data in the conventional manner for LSTM based models, where algorithms not solely consider attributes but additionally considering previous events or rows, it results in a deterioration in model performance. Consequently, our proposed model with specific modeling improved efficiency of the model proposed.

The aim was also to preserve simplicity for two reasons, faster learning and model building, and more transparent implementation unlike other complex solutions with good performance [4]. Imbalanced by nature, credit scoring data may result in degraded model performance. To address this problem, various techniques can be applied but the most common and the most effective in credit score is SMOTE [5] as an oversampling technique that generates synthetic instances of minority classes. Experimental results indicated that proposed stacked deep learning models are generally more competitive than other existing models, but utilization of SMOTE can lead to overfitting due to the small size of the dataset.

This paper is organized as follows: In this section, the credit scoring problem was shortly introduced and explained. Section two gives insight into previous relevant research in the field and reviews the related work. Section three describes the data structure, theoretical foundations, and techniques used in this study. The real-world implementation of the proposed methods is described in detail together with an explanation of the novelty approach to dataset modeling. Discussion and analysis of the achieved results are stated in section four. Finally, section five presents the study's conclusions and outlines recommendations for future research.

II. Background And Related Work

The field of credit scoring has a rich history, and its evolution continues to be driven by advancements in algorithmic techniques, processing capabilities, as well as memory and storage capacities. Despite the extensive literature on credit scoring, there is still a search for an optimal classifier, given the constant demand for improved performance and accuracy [6]. Currently, there is a growing trend in the credit scoring field towards the utilization of deep learning algorithms. One of the pioneering deep learning implementations in credit scoring involved the application of deep belief networks for corporate default prediction. [7]. Deep belief networks with Restricted Boltzmann Machines [8] were also utilized and credit risk classification was performed using an ensemble of deep MLP networks and deep convolutional neural networks [9]. Deep convolutional neural networks significantly outperformed DMLP [10]. Credit scoring models using the power of the CNN method and relief algorithm by converting data sets into images by bucketing features and mapping them into 2-D pixel matrix were also analyzed [11] but with the Chinese consumer finance company dataset, not the classical German or Australian Credit datasets. A 29-layer network with different machine learning algorithms (Deep Genetic Hierarchical Network of Learners) with limitations in extreme complexity and long-term training was proposed by [12]. Further, back-propagation artificial neural networks (BP-ANN) are applied and analyzed for credit scoring utilization [13].

According to the review [14], LSTM was not used as a technique for solving credit scoring problems, but it is evident that many papers that covered LSTM mostly belong to other financial fields, and some of the credit scoring applications were published after the period covered by the review. Although researchers like [15] and [16] examined the LSTM algorithm for solving credit scoring problems, it was still not applied to classical public German or Australian datasets but to a massive transactional dataset with millions of transactions which is again the sequence-based dataset. Another study [17] utilizes a single LSTM network but on peer-to-peer lending credit scoring which outperforms traditional statistical and ML models. DFNNs (A Dynamical Feedforward Neural Network) with the employment of a family of nonlinear neural network activation functions were also applied and analyzed [18]. Additional novel techniques were also analyzed and compared for German credit scoring models such as Step-wise multi-grained augmented gradient boosting decision trees [19], Dynamic 1-Nearest Neighbor [20] and Cost-sensitive Neural Network Ensemble [21].

To the best of the authors' knowledge, the use of Stacked LSTM and Stacked BiLSTM for credit scoring has not been widely investigated, except for a previous study [3] that transformed the attributes of each loan instance into a sequence of matrices using a fixed sliding window approach with a one-time step. Therefore, this paper proposes novel credit scoring models based on Stacked LSTM and Stacked BiLSTM architectures, and investigates their performance with the application of the SMOTE technique. SMOTE as a technique was applied together with LSTM [5] but Stacked LSTM and BiLSTM together with SMOTE were not used for credit scoring prediction by now.

Reinforcement learning and deep learning in general, can effectively forecast and detect difficult market trends compared to the traditional ML algorithms, with the significant advantage of high-level feature extraction properties and proficiency of the problem solver methods constructing more efficient framework considering crucial market constraints by integrating the prediction problem with the portfolio structure task [3] [22].

III. Proposed Deep Learning Model

A. Problem Description

It is evident that deep learning techniques can outperform not only statistical methods, but also traditional machine learning techniques. However, not all deep learning techniques are applicable for solving problems such as credits scoring since they are tailored to operate with time sequence datasets. For algorithms such as Stacked LSTM and BiLSTM, it was the challenge to shift perspective on how algorithms can exploit its possibilities without losing the predictive power by putting the classes of independent loan applicant into a sequence of events because they are not correlated nor connected by any means, since classes are only dependent on applicant's own attributes.

B. Preliminaries

Although primarily developed to deal with sequential data, the LSTM network is very effective in mining the interrelationships between credit data variables [5]. LSTM (Long Short-Term Memory) networks are a special type of RNN with weight updates and optimization methods that operate on the same principle but are structured to remember and predict based on short and long-term dependencies that are trained with time series data. It consists of numerous memory cells. LSTM units of LSTM networks are composed of three special multiple-cell gates that regulate information flow: input, output, and forget gate [23].

Stacked LSTM is a novel architecture where the output of a hidden layer will be fed as the input into the subsequent hidden layer and differs from classic LSTM in the number of those hidden layers [24]. Formulas 1-6 represent the form of the forward pass of the LSTM unit:

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \quad (1)$$
$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f) \quad (2)$$
$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \quad (3)$$
$$\tilde{c}_t = tanh(w_c[h_{t-1}, x_t] + b_c) \quad (4)$$
$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \quad (5)$$
$$h_t = o_t * tanh(c_t) \quad (6)$$

where i_t stands for input gate, f_t represents forget gate, o_t output gate, x_t - current timestamp, b_f - bias vector specific to the forget gate, h_{t-1} represents LSTM output of the previous timestamp, w – weights, σ - sigmoid function, tanh - hyperbolic tangent function, and * is the element-wise vector/matrix multiplication operator.

Input gate determines how much new information should be added to the cell state (1). The input gate takes the current input, typically combined with the previous output, and passes it through a sigmoid activation function σ. This function outputs values between 0 and 1, representing the relevance of the input (1). The equation (2) represents the computation of the forget gate activation at time step t. It takes as input the concatenation of the previous hidden state h_{t-1} and the current input x_t, linearly transforms it using weight matrix w_f and bias vector b_f, and applies the sigmoid activation function σ. The output gate determines how much of the updated cell state should be exposed as the hidden state at the time step t (3). Formula (4) represents a proposal for the new cell state at the current time step. The equation (5) shows updated cell state. Additionally, the current cell state, processed by the tanh function, is multiplied by the output gate output (6). The resulting values are the LSTM cell's output.

Stacked BiLSTM comprises at the same time characteristics of Bidirectional RNNs, and Stacked LSTM takes both context information by concatenating left and right summary vectors, thus performing better than unidirectional deep neural architecture. A BiLSTM (bidirectional LSTM) layer is exploited to capture spatial features and bidirectional temporal dependencies from historical data [24]. Models can be additionally built as a mixture of both, Stacked LSTM and BiLSTM layers. The number of layers can vary and although it is considered that architecture with more hidden layers can result in better performance, it should not have to be the case in practice [3].

C. Data Acquisition

In this paper the German credit dataset [25] accessible from the UCI website will be applied since it is used in most of the research in the credit scoring field. German dataset is naturally imbalanced with 300 instances of “bad” class and 700 of “good” class with an imbalance ratio of 2.33. The number of attributes is 20.

The model includes a transformation of the dataset into matrices to be used with Stacked LSTM and BiLSTM algorithms. LSTM-based algorithms are designed for handling sequences of time-dependent events while credit score data sets represent data instances with customer demographics, application forms, and transactional data from customer history sublimed into one record.

Sliding window chunks are not shaped to consider past events due to the nature of the credit score dataset, but with fixed sliding approach without overlapping of time series data where features themselves are treated as time series data and the next value to predict is a class itself.

To feed the neural network with an appropriate 3D array, the training dataset is transformed into shape (937, 1, 20). The 21st column represents the result set, where each of the 21 units is passed independently into the LSTM or BiLSTM layer. Each data point was a three-dimensional array with one time-step and 20 features. The rest of the 63 instances are divided into unseen test data that are not trained by our model but divided for testing purposes to evaluate the model. That number of 63 instances was chosen randomly because the dataset is small, but extraction was done after shuffling. Data is shaped by the way that features become a sequence of data with a sliding window of size one since we did not want to overlap due to the nature of credit scoring data where credit records are not dependent on each other considering a time perspective. The accuracy calculation is performed at the end of each epoch to monitor the model's performance over time. It helps assess how well the LSTM model is learning and making predictions as it iteratively updates its weights during training. The accuracy evaluation is conducted at the conclusion of each epoch to track the LSTM model's progression over the course of training. This assessment serves to gauge the model's proficiency in learning from the data and generating predictions while iteratively refining its weight parameters.

Due to the imbalanced nature of credit score dataset, SMOTE technique dataset is also included prior to training and building Stacked LSTM and BiLSTM neural networks within our proposed model.

D. Architecture of Proposed System

Selecting the optimal model for credit scoring requires considering not only metrics such as precision, accuracy, recall, and others, but also the complexity of implementation, development time, and interpretability of the model's decisions.

The deep learning models in this study are implemented in Python using the Keras Application Programming Interface, which runs on TensorFlow. While Keras and other open-source libraries provide a platform for building complex deep learning models, selecting the appropriate methods and hyperparameters is crucial to achieve superior performance. In addition to Keras, we employed libraries such as sklearn, pandas, matplotlib, and numpy to support the development and evaluation of our models. By carefully tuning the hyperparameters, even slight reductions in loss can lead to significant improvements in performance.

Stacked LSTM and BiLSTM are experimented in this research. Deep neural networks are designed as three-layer networks with 60 nodes for the input layer, and 60, 80, and 120 nodes for the first, second, and third layer respectively. Stacked LSTM and BiLSTM Architecture is shown in Fig. 1.

The number of units in the Dense layer is one since as the output we have information non-default/default (1/0). As an activation function, it is chosen Relu (Rectified Linear Unit), which is responsible for processing weighted inputs and helping to deliver an output. The epoch number is also a very important parameter that can lead to overfitting if it is too high or can lead to underfitting if too low. In our experiment, the number of epochs is set to 50. For the batch size it is used 32 and it defines the number of input-output pairs that are passed before updating internal model parameters. The dropout parameter in our experiment is set to 0.2. Other hyperparameters such as decay rate, learning rate, and momentum are set to default values.

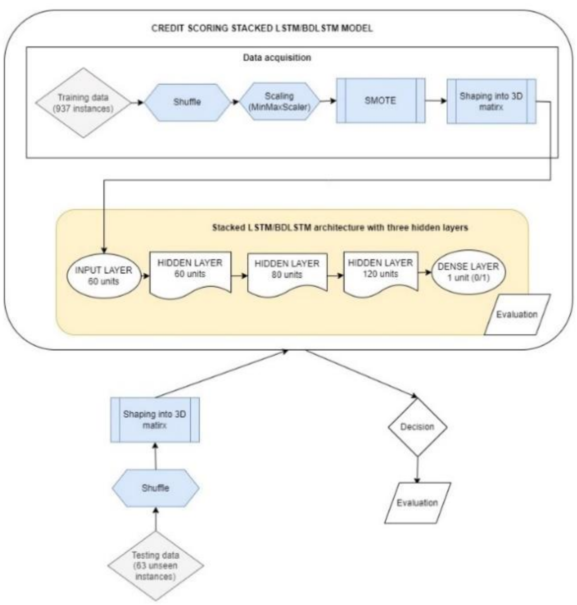

Fig.1. The three-layer Stacked LSTM and BiLSTM model.

IV. Results and Discussion

In this section, the performance of proposed Stacked LSTM and BLSTM models with three hidden layers is examined and compared. Prior to being passed to the next layer, the features undergo pooling and flattening operations. Effective data shaping, combined with appropriate hyperparameter tuning, is critical in maximizing a model's performance, as discussed in Section Three. To address class imbalance in the German credit dataset, SMOTE technique was utilized. The novel architectures were compared to other methods, demonstrating significant success through empirical evaluation. To evaluate the effectiveness of the proposed classifiers, standard performance metrics commonly used in the literature, such as Accuracy, Precision, Recall, and F1, were employed. Results on Stacked LSTM before and after the application of SMOTE technique are shown and compared in Table I.

TABLE I. German credit scoring model performance for Stacked LSTM

Metrics	German credit scoring model performance for Stacked LSTM	
	Stacked LSTM	***Stacked LSTM with SMOTE***
Accuracy (training)	**82.28%**	**88.09%**
Accuracy (testing)	**84.13%**	**79.36%**
Precision (macro avg)	79%	70%
Precision (wighted avg)	84%	78%
Recall (macro avg)	81%	74%
Recall (weighted avg)	83%	73%
F1 (macro avg)	80%	71%
F1 (weighted avg)	83%	74%

Credit scoring model performance for Stacked BiLSTM before and after the introduction of SMOTE is shown in Table II.

TABLE II. German credit scoring model performance for Stacked BDLSTM

Metrics	German credit scoring model performance for Stacked BDLSTM	
	Stacked LSTM	***Stacked LSTM with SMOTE***
Accuracy (training)	**87.19%**	**92.06%**
Accuracy (testing)	**88.89%**	**82.53%**
Precision (macro avg)	85%	73%
Precision (wighted avg)	87%	80%
Recall (macro avg)	83%	77%
Recall (weighted avg)	87%	76%
F1 (macro avg)	84%	74%
F1 (weighted avg)	87%	77%

The performances of different algorithms based on published results from other publications, compared to Stacked LSTM and Stacked BiLSTM application without oversampling of German credit, considering different measures are shown in Table III.

TABLE III. PERFORMANCE OF ML AND DL MODELS FOR THE GERMAN CREDIT DATASET

Algorithm	Metrics			
	Accuracy	*Precision*	*Recall*	*F1*
kNN	67.71%	73.74%	83.36%	78.01%
CART	64.71%	74.97%	74.64%	75.01%
NB	70.57%	80.89%	75.05%	77.77%
SVM	69.42%	70.43%	96.30%	81.25%
DSNN	75.90%	81.00%	82.00%	81.50%
CNN	76.00%	79.00%	80.00%	79.50%
Stacked LSTM	82.18%	84.00%	83.00%	83.00%
Stacked BDLSTM	87.19%	87.00%	87.00%	87.00%

The evaluation metrics employed for imbalanced datasets, namely the H-measure, Brier score, and G-mean, are presented in Table IV.
The results also demonstrate a degradation in performance upon the incorporation of SMOTE based on those measures.

TABLE IV. GERMAN CREDIT SCORING MODEL USING IMBALANCE DATASET MEASURES

Algorithm	Metrics for imbalanced datasets		
	H-measure	*Brier score*	*G-mean*
LSTM without SMOTE	0.847058824	0.206349206	0.695665593
LSTM with SMOTE	0.727272727	0.333333333	0.695665593
BiLSTM without SMOTE	0.876404494	0.174603175	0.791154805
BiLSTM with SMOTE	0.790123457	0.26984127	0.743697801

To adhere to the recommended generalization strategy, an unseen test dataset was employed to evaluate the model in the experimental setup. The results in Table I clearly demonstrate that the utilization of the SMOTE technique resulted in overfitting, as it improved the accuracy on the training data while causing a decline in the results on the test data.

SMOTE technique creates additional samples by taking the difference between the sample and the closest neighbor where the difference is multiplied by a random number between 0 and 1 and added to the feature vector [26]. In our case with SMOTE technique, additional 400 instances were generated; thus, minority and majority classes were equalized. Fig. 2. shows examples of underfitting, overfitting, and appropriate fitting [27].
In the case of applying Stacked LSTM and Stacked BiLSTM without SMOTE, appropriate fitting is achieved, but after SMOTE technique was introduced, we faced over-fitting according to the results.

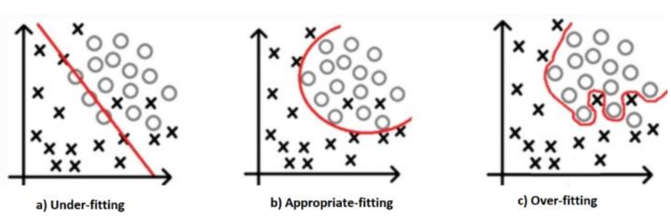

Fig.2. Illustration of three prediction model cases

However, by applying SMOTE techniques, we achieved results with the highest training accuracy of 88.09 % for Stacked LSTM and 92.06 % for Stacked BiLSTM and testing accuracy of 79.36 % and 82.53 % respectively.

V. CONCLUSION AND FUTURE RESEARCH

Despite the relatively small size of the German dataset, our experiments have demonstrated the effectiveness of the proposed techniques, which achieved high levels of accuracy on unseen datasets even without the need for oversampling.

Designing an adequate prediction model for German credit score is a challenging task due to the extremely small size of the dataset. Deep learning models require more data for training to avoid overfitting [28]. In practical applications, credit data is often readily available in larger quantities, making it more suitable for the implementation of deep learning algorithms. As a result, these algorithms can achieve significantly higher precision and overall performance. On the other hand, LSTM and BiLSTM as RNN-based models are designed to predict the next value of time series while credit scoring rows consist of aggregated attributes that usually contain historical data related to the personal, financial, employment information, previous behavior, and loan information labels. The innovation of this work was to adapt credit scoring data to be successfully applied to deep neural architectures which operate with a 3D array of time sequence data but on such a way that rows between themselves are should not be considered as dependent since they are not dependent in practice. Method overperformed other, even more complicated algorithms. In previous studies, details on data shaping in most cases were not provided, so that gap is also closed.

Another limitation for DL based models is the fact that instances are also not correlated as it is case with stock prediction or monthly movement of bank reserves.

Although traditional machine learning algorithms have been commonly used for credit risk prediction, our research has demonstrated the superiority of the proposed deep learning model in improving forecast accuracy and default detection. However, the application of SMOTE technique in our model led to a decline in performance on the test dataset, resulting in overfitting. Models proposed achieved distinguished results and still preserved its simplicity. Moreover, the use of a Bidirectional LSTM (BiLSTM) architecture allows for the incorporation of both left and right context information, resulting in superior predictive performance when compared to unidirectional LSTM models.

In conclusion, deep learning represents a powerful solution for credit risk assessment, offering superior performance when compared to traditional methods. However, due to regulatory requirements imposed by banking agencies, traditional methods still dominate the industry, as they provide a more transparent explanation of decision-making. Nevertheless, we believe that with continued research in the area of interpretable AI, it will be possible to address these concerns and facilitate the widespread implementation of deep learning algorithms for credit risk assessment in the future.

REFERENCES

[1] A. Gicić and A. Subasi, "Credit scoring for a microcredit data set using the synthetic," *Expert Systems,* vol. 36, no. 6, 2019.

[2] J. N. Crook, D. B. Edelman and T. C. Lyn, "Recent developments in consumer credit risk assessment," *European Journal of Operational Research,* vol. 183, no. 3, pp. 1447-1465, 2007.

[3] A. Gicic, D. Đonko and A. Subasi, "Intelligent Credit Scoring using Deep Learning," *Concurrency and Computation: Practice and Experience,* 2023.

[4] P. Pławiak, A. Moloud and A. U. Rajendra, "Application of new deep genetic cascade ensemble of SVM classifiers to predict the Australian credit scoring," *Applied Soft Computing,* vol. 84, no. 105740, 2019.

[5] F. Shen, X. Zhao, G. Kou and F. E. Alsaadi, "A new deep learning ensemble credit risk evaluation model with an improved synthetic minority oversampling technique," *Applied Soft Computing,* vol. 98, no. 1, p. 106852, 2021.

[6] M. Ala'raj and M. F. Abbod, "A new hybrid ensemble credit scoring model based on classifiers consensus system approach," *Expert Systems with Applications,* vol. 64, pp. 36-55, 2016.

[7] S.-H. Yeh, C.-J. Wang and M.-F. Tsai, "Deep belief networks for predicting corporate defaults," in *24th Wireless and Optical Communication Conference (WOCC)*, Taipei, Taiwan, 2015.

[8] C. Luo, D. Wu and D. Wu, "A deep learning approach for credit scoring using credit default swaps," *Engineering Applications of Artificial Intelligence,* vol. 65, pp. 465-470, 2017.

[9] Y. Li, X. Lin, X. Wang, F. Shen and Z. Gong, "Credit Risk Assessment Algorithm Using Deep Neural Networks with Clustering and Merging," in *201 13th International Conference on Computational Intelligence and Security (CIS)*, Hong Kong, China, 2017.

[10] V.-E. Neagoe, A.-D. Ciotec and G.-S. Cucu, "Deep convolutional neural networks versus multilayer perceptron," in *International Conference on Communications (COMM)*, 2018.

[11] 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)*, Chengdu, China, 2018.

[12] P. Pławiak, M. Abdar, J. Pławiak, V. Makarenkov and R. Acharya, "DGHNL: A New Deep Genetic Hierarchical Network of Learners for Prediction of Credit Scoring," *Information Sciences,* pp. 401-418, 2020.

[13] R. Zhang and Q. Zhiyi, "Optimizing hyper-parameters of neural networks with swarm intelligence: A novel framework for credit scoring," *PLoS One,* vol. 15, no. 6, 2020.

[14] O. B. Sezer, M. U. Gudelek and A. M. Ozbayoglu, "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019," *Applied Soft Computing Journal,* vol. 90, no. 106181, 2020.

[15] D. Babaev, M. Savchenko, A. Tuzhilin and D. Umerenkov, "E.T.-RNN: Applying Deep Learning to Credit Loan Applications," in *KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019.

[16] C. Wang, D. Han 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.

[17] Y. Zhang, D. Wang, Y. Chen, H. Shang and Q. Tian, *Intelligent Computing Theories and Application,* 2017.

[18] S. Mahajan, A. Nayyar, A. Raina, S. J. Singh, A. Vashishtha and A. K. Pandit, "A Gaussian process-based approach toward credit risk modeling using stationary activations," *Concurrency and Computation Practice and Experience,* vol. 34, no. 5, 2021.

[19] W. Liu, H. Fan and M. Xia, "Step-wise multi-grained augmented gradient," *Engineering Applications of Artificial Intelligence,* vol. 97, 2021.

[20] O. Camacho-Urriolagoitia, I. López-Yáñez, Y. Villuendas-Rey, O. Camacho-Nieto and C. Yáñez-Márquez, "Dynamic Nearest Neighbor: An Improved Machine Learning Classifier and Its Application in Finances," *Applied Sciences,* vol. 11, no. 19, p. 8884, 2021.

[21] W. Yotsawat, P. Wattuya and A. Srivihok, "A Novel Method for Credit Scoring Based on Cost-Sensitive Neural Network Ensemble," *IEEE Access,* vol. 9, pp. 78521-78537, 2021.

[22] V. K. Singh, S.-S. Chen, M. Singhania and B. Nanavati, "How are reinforcement learning and deep learning algorithms used for big data based decision making in financial industries-A review and research agenda," *International Journal of Information Management,* vol. 2, no. 2, 2022.

[23] S. Hochreiter and J. Schmidhuber, "Long Short-term Memory," *Neural computation,* vol. 9, p. 1735– 1780, 1997.

[24] Z. Cui, R. Ke and Z. Pu, "Stacked Bidirectional and Unidirectional LSTM Recurrent Neural Network for Network-wide Traffic Speed Prediction," *Transportation Research Part C Emerging ,* 2018.

[25] H. Hofmann, "UCI Machine Learning Repository: Statlog (German Credit Data) Data Set," *Institut fur Statistik und "Okonometrie Universit" at Hamburg,* 1994.

[26] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *The Journal of Artificial Intelligence Research,* vol. 16, p. 321–357, 2002.

[27] O. A. M. López and A. . M. López, "Overfitting, Model Tuning, and Evaluation of Prediction Performance," in *In: Multivariate Statistical Machine Learning Methods for Genomic Prediction*, Springer, 2022.

[28] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data volume,* vol. 6, no. 60, 2019.