Credit Score Prediction using Genetic Algorithm-LSTM Technique

Juliana Adisa, Samuel Ojo, Pius Owolawi, Agnieta Pretorius
Department of Computer Systems and Engineering
Tshwane University of Technology
Pretoria, South Africa
219940939@tut4life.ac.za, 208322842@tut4life.ac.za
owolawipa@tut.ac.za, pretoriusab1@tut.ac.za

Sunday O. Ojo
Department of Information Technology
Durban University of Technology
Durban, South Africa
sundayO1@dut.ac.za

2022 Conference on Information Communications Technology and Society (ICTAS) | 978-1-6654-4019-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICTAS53252.2022.9744714

Abstract—In data mining, the goal of prediction is to develop a more effective model that can provide accurate results. Prior literature has studied different classification techniques and found that combining multiple classifiers into ensembles outperformed most single classifier approaches. The performance of an ensemble classifier can be affected by some factors. How to determine the best classification technique? Which combination method to employ? This paper applies Long Short-Term Memory (LSTM), one of the most advanced deep learning algorithms which are inherently appropriate for the financial domain but rarely applied to credit scoring prediction. The research presents an optimization approach to determine the optimal parameters for a deep learning algorithm. The LSTM parameters are determined using an optimization algorithm. The LSTM parameters include epochs, batch size, number of neurons, learning rate and dropout. The results show that the optimized LSTM model outperforms both single classifiers and ensemble models.

Keywords— Long Short-Term Memory; genetic algorithm; credit scoring; credit predicgtion.

I. INTRODUCTION

Financial institutions face the important but challenging task of developing an effective credit prediction model. A credit score or rating is applied to determine whether individuals are good or bad candidates for loans. Credit rating is primarily concerned with the identification of good or bad applicants. The failure and inability of the prediction models to provide a certain level of accuracy would result in incorrect decisions and lead to severe problems. The financial organizations employed a credit scoring model to determine credit extension to a prospective customer. All stakeholders have relied on credit ratings to assess the risk premium and the marketability of bonds.

Several studies have demonstrated that neural networks, a machine learning technique, perform better in prediction accuracy and error than conventional statistical methods such as logistic regression [1, 2]. According to Lin et al., classifier ensembles are far more effective than single classifiers [3, 4]. However, most ensemble classifiers constructed for financial prediction are incredibly narrow classifiers. In addition, Dietterich [5] concluded that the ensemble classifiers performances rely on either bagging or boosting, which are combination methods. These combination methods are usually domain-dependent in bagging and boosting [6].

Recently, deep learning techniques have become increasingly popular for classification prediction, and their deployment to different fields has increased. In artificial neural networks (ANNs), deep learning models are characterized by having more than one hidden layer between the input layer and the output layer. Deep learning methods include recurrent neural networks (RNN), convolutional neural networks (CNN), and deep belief networks (DBN). Long short-term memory (LSTM) is a type of RNN that employs feedback connections within the network to account for the temporal effects of past significant events [7]. Hence, LSTM can represent temporal sequence data, and it is beneficial for tasks like financial predictions, speech recognition, and natural language processing. Conventional neural networks lack consideration for the temporal effects of past events. Hence, ANNs are weak and cannot deal with sequential data effectively. This study will apply LSTM for the credit scoring prediction to overcome the limitations of the conventional ANN. In addition, the paper presents a method developed by employing a genetic algorithm to optimize a long-short term memory (LSTM) for credit scoring. The proposed method is validat+ed using existing methods. These methods include artificial neural networks, support vector machines, decision trees. The proposed model demonstrated to be the best alternative to these methods. An overview of related research works is presented in Section II. Section III presents the methodology of the models. In Section IV, we describe the experimental framework. Section V discusses the results and the conclusion in Section VI.

II. LITERATURE REVIEW

There are two main branches of credit prediction models. First, statistical methods are sub-categorized as supervised (binary classification problem) and unsupervised (outliers' detection). The most common statistical models for credit prediction in the review are linear discriminant analysis (LDA) and logistic regression (LR) [8]. There was, however, a report that LDA and logistic regression are incapable of providing sufficient credit scoring accuracy if the underlying relationship between variables is linear [9, 10]. The second group use artificial intelligence (AI) methods. AI is divided into machine learning and deep learning techniques. In 1990, some studies showed effectiveness in credit scoring prediction [7]. In 1996,

978-1-6654-4019-6/22/$31.00 ©2022 IEEE

Desai et al. [11] developed credit scoring models by employing neural networks (NN), LDA and LR techniques, and it was concluded that both NN and LR had the best performance [11]. Table 1 shows the deep learning techniques applied for credit prediction using a specific dataset.

TABLE I. DEEP LEARNING APPLICATION FOR CREDIT SCORING

Ref.	Dataset	Models
[12]	The XR 14 CDS contracts	DBN+RBM
[13]	Japanese and German credit data	SVM + DBN
[14]	Kaggle credit data	DMLP
[15]	German and Australian credit data	GP + AE as Boosted DMLP
[16]	German and Australian credit dataset	DCNN, DMLP
[17]	Chinese consumer credit data	CNN + Relief
[18]	UCI Credit dataset	Rectifier, Tanh, Maxout DL

III. METHODOLOGY

We applied a classification algorithm called long short-term memory for credit card prediction. The optimization algorithm will optimize the learning parameters of the LSTM model to get an optimum result: the number of LSTM neurons, epochs, batch size, learning rate, and dropout. However, it is impossible to find an optimal set of space parameters because of time and computation constraints. In previous research, researchers depended on personal experience to determine these control parameters. Although LSTM networks are significant, only a limited amount of research has been conducted on optimal parameters. This research work introduced a genetic algorithm to find the optimal parameters for the LSTM model.

A. Long Short Term Memory Networks (LSTM)

Long Term Memory networks (LSTMs) solve the problem of long-term dependency. LSTMs are typically used to model sequence learning; it contains four layers of repeating units [19, 20]. The LSTM network has four gates constructed from sigmoid functions and pointwise multiplication operations shown in Fig 1. Information passes through these gates, and their equations are provided by [21]:

1) Step 1: Forget gate f^t identify the information that is not required and needs to be removed from the cell state.

$$f^t = \sigma\left(W_f[o^{t-1}, x^t] + b_f\right) \qquad (1)$$

2) Step 2: Input gate i^t determines what new information needs to be stored in the cells state

$$i^t = \sigma(W_i[o^{t-1}, x^t] + b_i) \qquad (2)$$

$$\hat{C}^t = tanh(W_c[o^{t-1}, x^t] + b_c) \qquad (3)$$

3) Step 3: Update the old cell state C^{t-1}*into a new cell state* C^t

$$c^t = f^t * c^{t-1} + i^t * \hat{C}^t \qquad (4)$$

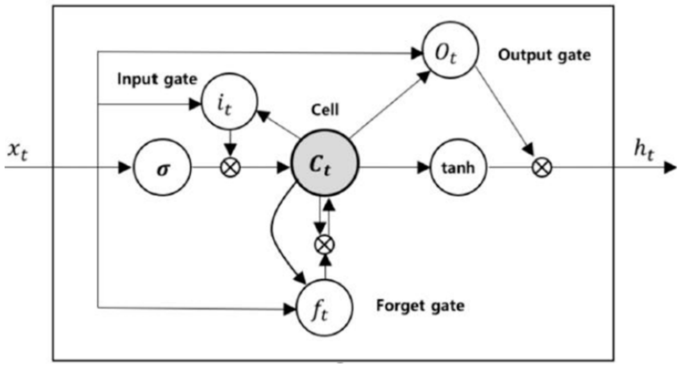

Fig. 1. LSTM Architecture

4) Step 4: Output gate; decide the output part

$$h^t = \sigma(W_h[o^{t-1}, x^t] + b_h) \qquad (5)$$

$$o^t = h^t * tanh(c^t) \qquad (6)$$

$$y^t = \sigma(W^{out} o^t + b^{out}) \qquad (7)$$

where W_f, W_i, W^{out}, represent the weight matrices at the forget gate, input gate and output gate respectively. The b_f, b_i, and b^{out}, are the bias for the respective gates. Also, x^t and o^{t-1} denote the input at time t and output at time t-1 respectively.

B. Genetic Algorithm

A genetic algorithm (GA) is an optimization technique used to generate solutions for problems. GA is part of the bigger group of algorithms called evolutionary algorithms, and they employ natural selection to approximate solutions for a given problem. GA used a population of possible solutions. Therefore, they are different from the conventional non-linear optimization models. Selection, crossover and mutation are the operators of GA.

1) Fitness function is a method of determining the quality of a solution. The algorithm calculates the fitness function for each individual, and it selects the solution with the highest fitness value from the population. Fitness functions determine the next generation of solutions based on the fitness value. Fitness functions are crucial to the overall performance of an algorithm. An algorithm with a poor fitness function may fail to find a solution.

2) Crossover is the process of creating new individuals by the selection of the parent's genome. The genome of the selected parents is cut off at a random point, and the switching of the genome of one parent with the genome of the other parent takes place. This crossover process is referred to as a single point crossover function, and it produces two new solutions for the next generation. This process is repeated till there is no specimen in the next generation.

3) Mutation helps to discover new solutions that were not possible during the previous steps. Mutation is the changing of bits of the genome with a certain probability. It means replacing the genome with randomized value. A new individual is created with just a single parent. Mutation helps with the following: it increases the diversity of the genetic population; it prevents the evolutionary process from becoming stuck in a local minimum because it prevents individuals from having two similar genomes.

The fitness function of each individual is calculated using the MSE in this study. As given in (8), y_i *and* $\tilde{y}$ represent the predicted and observed values for the credit scoring, respectively. The total number of instances is denoted by N.

$$f = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \tilde{y}_i)^2}{N}} \quad (8)$$

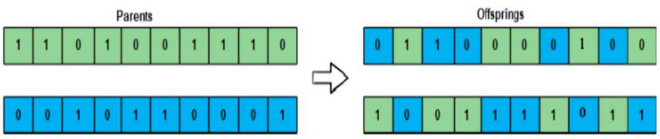

Fig. 2. Figure 1Genetic Algorithm Crossover.

C. Artificial Neural Networks (ANNs)

ANN is used to predict and process large datasets. It provides meaningful information from a large volume of data [22]. A neural network consists of neurons, connectivity patterns, propagation rules, learning rules, activation functions, and transfer functions. In ANNs, errors are minimized by adjusting the weights and subjecting the networks to backpropagation.

D. Support Vector Machines

Support Vector Machine (SVM) is a non-discriminatory learning model developed to separate training vectors into different categories using the hyperplane. Hyperplanes are boundary restrictions used to separate data patterns. The equation that expresses the hyperplane function as defined by [23],

$$f(x) = \langle w|x\rangle + b = \sum_{i=1}^{N}(w_i x_i) + b = 0 \quad (9)$$

E. Decision Tree

A decision tree is a tool used to classify or predict things based on a flowchart-like structure. It employed a "divide-and-conquer" approach to divide the data into leaves [24].

IV. Research Methodology

A. Data collection and preprocessing

The dataset used in this research is the Australian data downloaded from the UCI machine learning repository (https://archive.ics.uci.edu/ml/datasets/Statlog+(Australian+Credit+Approval)). The credit card data has 690 total instances, 14 attributes given in Table II. The attributes were renamed for privacy reasons. We employed the Australian data because it is balanced data, and there is no need to deal with any data skewness.

TABLE II. DATASET INFORMATION

Name of Data	Total instances	No. of attributes	Bad (Bankrupt) /Good (Non-Bankrupt)
Australian credit data	690	14	307/383

B. Experimental Setup

The Australian dataset was collected, preprocessed to remove outliers, and data normalized. We applied three binary classification algorithms, which are single classifiers. Note that the single classifiers are decision trees (DT), the multi-layer perceptron of artificial neural networks (MLP) and support vector machines (SVM) which are the baseline classifiers for this work. Ensemble models were developed from single classifiers using a fixed number of classifiers (10). The LSTM model was developed using the learning parameters in Table III.

We employed a genetic algorithm to optimize the LSTM network in the last section of experimental setups. Therefore, the hybrid LSTM-GA model was developed. The optimization algorithm was employed to investigate and find the optimal learning parameters for the LSTM model. The GA algorithm was used to determine the best epoch number, batch size, the total number of LSTM units, learning rate, and drop out for the LSTM architecture. Fig. 3 shows the flowchart of the framework for the improved credit prediction architecture. In addition, the LSTM model was optimized using a genetic algorithm with a learning parameter in Table IV.

TABLE III. LSTM LEARNING PARAMETERS

LSTM	Parameters
Batch size	100
Relu Gate Activation Function	Activation Sigmoid
Output Layer Activation Function	Activation Softmax
Number of epochs	10
Loss function	LossMCXENT

TABLE IV. GENETIC ALGORITHM LEARNING PARAMETERS

Parameter name	Parameter values
Chromosome size	10
Population size	50
Crossover rate	0.6
Mutation rate	0.5
Iteration	20

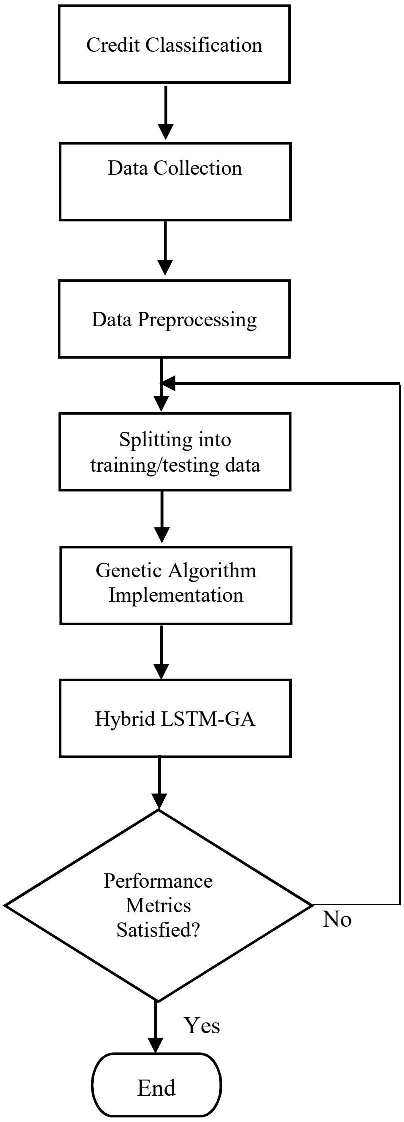

Fig. 3. An improved LSTM - GA credit prediction model.

C. *Performance Metrics*

In a classification model, there are four ways to classify an outcome: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). The evaluation metrics used the application of a confusion matrix. The metrics are Accuracy, Precision, Recall, and F1-Score, AUC, Computation time, Kappa statistics.

1) Accuracy can be measured as the percentage of predictions that are classified correctly, and it is given as:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

2) Root mean square error measure: First, it determine the difference between the predicted and actual values. Then the square root of the difference will be calculated.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \tilde{y}_i)^2}{N}}$$

where y_i and $\tilde{y}_i$ represent the actual and predicted value, respectively. The total number of instances is denoted by N.

3) Computation time measures the speed and the rate at which it converges.

4) A recall is defined as the ratio of the correctly classified positive instances to the total number of positive instances.

$$Recall = \frac{TP}{TP + FN}$$

5) Precision measures exactness. It estimates the number of true positives from the entire set of instances classified as true positives.

$$Precision = \frac{TP}{TP + FP}$$

6) ROC (AUC curves) indicates the degree of class separation.

7) Kappa statistics K is a measure of how closely two individuals agree that is not the result of chance [25].

$$K = \frac{P_o - P_e}{1 - P_e}$$

where P_o is the probability of agreement observed, while P_e is the probability of agreement expected by chance. Table V shows the kappa value with the agreement as given by [25].

8) F1-Score is a measure that represents the average of precision and recall.

V. Experimental Results and Analysis

This section presents experimental results for the proposed framework and discusses them. This research work was implemented using Python packages on the Google Collaboratory platform.

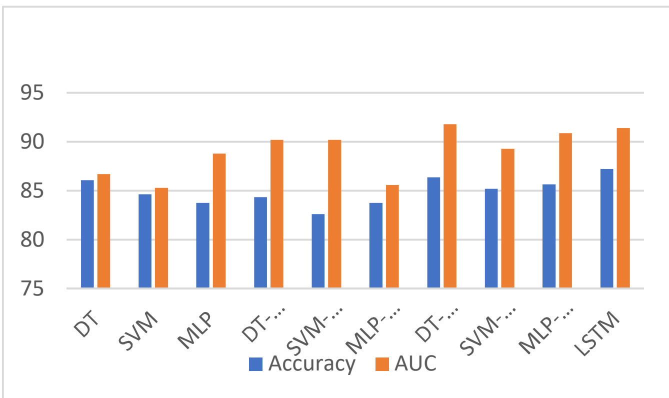

Fig. 4. Baseline models, Ensembles comparison to LSTM.

TABLE V. BASELINE , ENSEMBLE AND LSTM MODEL COMPARISON OF RESULTS

Models	Accuracy	RMSE	AUC	Computation time	Precision	Recall	F-Measure	Kappa
DT	86.09	0.3386	0.867	0.08	0.861	0.861	0.861	0.7183
SVM	84.64	0.3919	0.853	0.43	0.857	0.846	0.847	0.6941
MLP	83.77	0.3763	0.888	4.75	0.861	0.843	0.852	0.6722
DT-boosting	84.35	0.3696	0.902	0.17	0.853	0.867	0.860	0.6825
SVM-boosting	82.61	0.3459	0.902	0.94	0.854	0.828	0.841	0.6493
MLP-boosting	83.77	0.3868	0.856	13.26	0.861	0.843	0.852	0.6722
DT-Bagging	86.38	0.3215	0.918	0.06	0.881	0.872	0.877	0.7245
SVM-bagging	85.22	0.3575	0.893	0.9	0.863	0.852	0.853	0.7059
MLP-bagging	85.65	0.3363	0.909	45.81	0.876	0.864	0.870	0.71
LSTM	87.22	0.3366	0.914	358.42	0.854	0.885	0.869	0.7998

TABLE VI. IMPROVED LSTM-GA RESULT

	Epochs	Batch Size	No. of Neurons	Learning rate	Dropout	Loss	Accuracy
LSTM+GA	310	360	180	0.1261	0.3962	0.12008	89.27%

A. Baseline, Enemble and LSTM models Result

The 10-fold cross-validation was employed for the training/testing ratio. Table V shows the performance of the classifiers based on the evaluation metrics such as accuracy, AUC, precision, kappa, RMSE, precision, F-measure, recall, as explained in section IV. The computation time (seconds) measures the time duration for classifiers to complete the task. Table V shows the results is for the baseline models, ensemble models and the LSTM model. The LSTM model performs better in terms of accuracy (87.22%), and DT bagging is the best model when the AUC metric is employed (0.918). LSTM (accuracy) and DT bagging (AUC) are the best compared to the baseline and ensemble models. The computational time for the models is short, ranging from a few seconds to minutes. While it takes the LSTM model hours for completion, other models finish computation within a few seconds. Therefore, the LSTM model with a kappa value of 0.7998 has a substantial agreement with the result. The accuracy and the AUC for the baseline models, ensembles and LSTM were compared and given in Fig. 4.

B. Improved LSTM-GA Model

The genetic algorithm (GA) was applied to find the optimal architectural factors for the LSTM model. Hence, the LSTM-GA model was developed. There are 20 iterations for the LSTM-GA model, and there are 14 total attributes for the Australian dataset. The LSTM-GA was applied to the 14 features, and the best optimal result from the 20 iterations is given in Table VI and the corresponding graphs in Fig. 5 and Fig. 6. The hybrid LSTM-GA outperforms the ordinary LSTM model with an accuracy of 89.27%.

VI. CONCLUSION

In this study, we have investigated the performance of single classifiers and compared them to their corresponding ensemble models and hybrid models. We further investigate the performance of a long short-term memory (LSTM) model and introduce a genetic algorithm to optimize and generate the optimal hyperparameters for the LSTM model.

A genetic algorithm was employed as an optimization scheme to find the optimal parameters for the LSTM model. The result of the optimized model shows a better prediction performance and an improved loss than an ordinary LSTM model. The optimized model, therefore, outperforms the baseline and the ensemble models. Their confusion matrix shows that the LSTM model has 14.78% of its elements classified as either Type 1 or Type 11 error, while the optimized LSTM model has 13.76% of its features classified as error. The results clearly show that the optimized LSTM has better performance. Overall, the hybrid LSTM perform better than all models employed in this research.

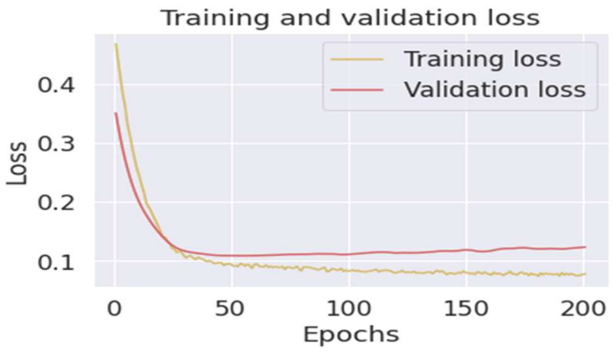

Fig. 5. Optimized LSTM-GA Loss. The loss graph shows a good fit, and the training loss is below the validation loss. Adding dropout will cause the model to overfit because the size of the Australian dataset.

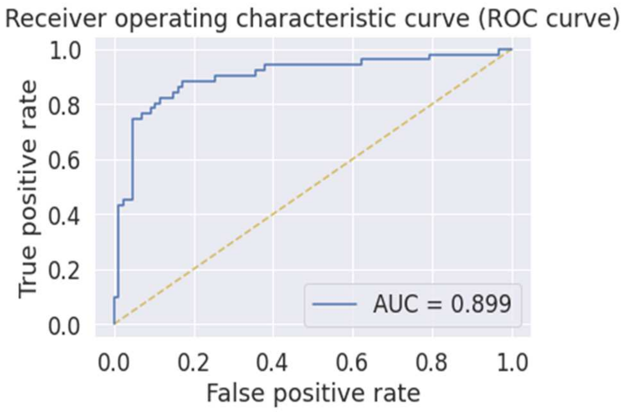

Fig. 6. LSTM-GA ROC Curve. The Area under the curve (AUC) 0.899 means that the model is 89% sure to distinguish between the dataset classes.

Acknowledgment

The authors acknowledge the Tshwane University of Technology for providing the facilities used by the authors and the National Research Foundation (NRF) under grant number 129088 in support of this study.

References

[1] C. F. Tsai, “Financial decision support using neural networks and support vector machines,” Expert Systems, vol. 25, no. 4, pp. 380-393, 2008.

[2] C.-F. Tsai, and J.-W. Wu, “Using neural network ensembles for bankruptcy prediction and credit scoring,” Expert systems with applications, vol. 34, no. 4, pp. 2639-2649, 2008.

[3] W.-Y. Lin, Y.-H. Hu, and C.-F. Tsai, “Machine learning in financial crisis prediction: a survey,” IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 4, pp. 421-436, 2011.

[4] J. A. Adisa, S. O. Ojo, P. A. Owolawi, and A. B. Pretorius, "Financial Distress Prediction: Principle Component Analysis and Artificial Neural Networks," in 2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC), 2019, pp. 1-6: IEEE.

[5] T. G. Dietterich, “An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization,” Machine learning, vol. 40, no. 2, pp. 139-157, 2000.

[6] C.-F. Tsai, Y.-F. Hsu, and D. C. Yen, “A comparative study of classifier ensembles for bankruptcy prediction,” Applied Soft Computing, vol. 24, pp. 977-984, 2014.

[7] A. M. Ozbayoglu, M. U. Gudelek, and O. B. Sezer, “Deep learning for financial applications: A survey,” Applied Soft Computing, vol. 93, pp. 106384, 2020.

[8] S.-T. Luo, B.-W. Cheng, and C.-H. Hsieh, “Prediction model building with clustering-launched classification and support vector machines in credit scoring,” Expert Systems with Applications, vol. 36, no. 4, pp. 7562-7566, 2009.

[9] L. C. Thomas, “A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers,” International journal of forecasting, vol. 16, no. 2, pp. 149-172, 2000.

[10] D. West, “Neural network credit scoring models,” Computers & Operations Research, vol. 27, no. 11-12, pp. 1131-1152, 2000.

[11] V. S. Desai, J. N. Crook, and G. A. Overstreet Jr, “A comparison of neural networks and linear scoring models in the credit union environment,” European journal of operational research, vol. 95, no. 1, pp. 24-37, 1996.

[12] 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.

[13] L. Yu, R. Zhou, L. Tang, and R. Chen, “A DBN-based resampling SVM ensemble learning paradigm for credit classification with imbalanced data,” Applied Soft Computing, vol. 69, pp. 192-202, 2018.

[14] Y. Li, X. Lin, X. Wang, F. Shen, and Z. Gong, "Credit risk assessment algorithm using deep neural networks with clustering and merging," in 2017 13th International Conference on Computational Intelligence and Security (CIS), 2017, pp. 173-176: IEEE.

[15] K. Tran, T. Duong, and Q. Ho, "Credit scoring model: A combination of genetic programming and deep learning," in 2016 Future Technologies Conference (FTC), 2016, pp. 145-149: IEEE.

[16] V.-E. Neagoe, A.-D. Ciotec, and G.-S. Cucu, "Deep convolutional neural networks versus multilayer perceptron for financial prediction," in 2018 International Conference on Communications (COMM), 2018, pp. 201-206: IEEE.

[17] 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), 2018, pp. 205-208: IEEE.

[18] A. Niimi, "Deep learning for credit card data analysis," in 2015 World congress on internet security (WorldCIS), 2015, pp. 73-77: IEEE.

[19] S. O. Ojo, P. A. Owolawi, M. Mphahlele, and J. A. Adisa, "Stock market behaviour prediction using stacked LSTM networks," in 2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC), 2019, pp. 1-5: IEEE.

[20] S. O. Ojo, J. A. Adisa, P. A. Owolawi, C. T. Du, and M. Mphahlele, "Stock Market Behaviour Prediction using Long Short-Term Memory Network and Gated Recurrent Unit," in 2020 International Conference on Computational Science and Computational Intelligence (CSCI), 2020, pp. 615-621: IEEE.

[21] G. Gilardoni, “Recurrent neural network models for financial distress prediction,” 2017.

[22] J. H. Min and C. Jeong, "A binary classification method for bankruptcy prediction," *Expert Systems with Applications,* vol. 36, no. 3, pp. 5256-5263, 2009.

[23] L. Hardinata and B. Warsito, "Bankruptcy prediction based on financial ratios using Jordan Recurrent Neural Networks: a case study in Polish companies," in *Journal of Physics: Conference Series*, 2018, vol. 1025, no. 1, p. 012098: IOP Publishing.

[24] T.-N. Chou, "An Explainable Hybrid Model for Bankruptcy Prediction Based on the Decision Tree and Deep Neural Network," in *2019 IEEE 2nd International Conference on Knowledge Innovation and Invention (ICKII)*, 2019, pp. 122-125: IEEE.

[25] J. R. Landis, and G. G. Koch, “The measurement of observer agreement for categorical data,” biometrics, pp. 159-174, 1977.