International Journal of Emerging Research in Engineering and Technology
Pearl Blue Research Group| Volume 7 Issue 1 PP 38-46, 2026
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.IJERET-V7I1P106r

Original Article

Efficient Deep Learning Models for Accurate Default Loan Prediction in Credit Risk Management

Deepak Reddy Suram
Senior Software Engineer & Cloud Data Architect H&R Block, Inc.

Received On: 27/11/2025 Revised On: 30/12/2025 Accepted On: 06/01/2026 Published On: 13/01/2026

Abstract *- The increasing complexity of lending environments has enhanced the importance of complex credit risk measurement procedures that have the capacity of analyzing massive, nonlinear, and highly imbalanced financial information. To solve this, a deep-learning method was used to forecast credit defaults with the complete Lending Club data that contains a variety of demographic, behavioral, and financial attributes of borrowers. Data preparation was reliable due to rigorous preprocessing, includes using SMOTE to address missing data, outliers, and class imbalance. A comparison between ANN and ResNet, FKNN, and AdaBoost was created and compared. The results of studies show that ANN outperforms all other baseline models, achieving 98.92% accuracy, 98.08% precision, 99.2% recall, 99.1% specificity, 98.63% F1-score, and 0.99 AUC respectively. Additionally, the ROC curve and confusion matrix data confirm the good class-separation capacity and low rate of misclassification of the model. These results indicate the ANN usefulness in mimicking intricate borrower trends and improving predictive validity. On the whole, the findings demonstrate the importance of deep-learning methods to enhance credit risk evaluation and allow more effective, data-driven decision making in the framework of current automated underwriting systems.*

Keywords *- Credit Default Prediction, Deep Learning, Financial Risk Assessment, Loan Default Modeling, Machine Learning, Predictive Analytics.*

1. Introduction

The financial stability of any economy is pegged on the management of credit risk, since the proper forecasting of defaults on loans directly affects the profitability and risk exposure of financial institutions [1]. Banks are especially important in developing economies, where borrowers have limited access to capital markets. In such situations the stability and efficiency of banks does not only lead to financial inclusion but also economic growth. Although banks are fundamental, they are prone to numerous risks and can be grouped into financial, operational, and strategic risks, where one of the most important aspects of financial risk is credit risk [2][3]. Credit risk is the possibility that an investment's real yield or loan might not turn out as per the expectations because of inability or unwillingness of the persons who took the loans to fulfil their obligations [4][5]. Such factors as the weak institutional capacity, ineffective credit policies, fluctuating interest rates, ineffective management, irresponsible lending policies, and ineffective regulation also play a major role in the rate of loan defaults and bad loans, which present significant problems to financial stability [6].

In loan default predicting, the credit risk assessment techniques like manual audits and statistical means have been greatly employed. Such methods as they are have provided some background details, but in their nature they are weak in terms of representing the non-linear relationship that is intricate in financial data [7]. The techniques of the static modelling fail to deal with the dynamism of the financial environment, which is constantly faced by a sudden change in policies or unexpected market shocks that can radically alter the behavior of borrowers [8]. Also, the traditional models are not always applicable to high-dimensional and uneven data, limiting their ability to predict. Hence, the default probability calculation, which is an essential measure in the credit risk management context, is an undertaking that is problematic and more sophisticated and automated methods of modeling are required.

The past few years have seen machine learning as an efficient, reliable, and flexible financial risk prediction tool. Such methods as neural networks, the use of advanced data mining techniques and ensemble methods have been more applied in predicting the risk of default in loans [9][10]. These methods are dramatically more effective in predictive accuracy than traditional logistic regression or other two-stage assessment models do because they incorporate complex patterns in the information of borrowers [11]. However, conventional machine learning models tend to be limited by the necessity to use hand-constructed feature engineering, which may restrict the richness of the information extracted, as well as the ability to make qualitative contributions to prediction accuracy.

Potential remedies for these flaws may be found in DL networks like autoencoders, CNN, LSTM, and ANN. DL models achieve this by automatically learning hierarchical representations of high-dimensional data, which are not possible to represent by other methods [12][13]. They are especially applicable in loan default prediction because of their capability to efficiently handle large volumes of financial information and predict it with a high degree of accuracy. The use of these models can thus improve the decision-making process in credit risk management and minimize non-performing loans, which can help the entire financial institution to remain stable.

1.1. Motivation and Contribution of the Study

This study is motivated by the increasing demand to develop intelligent and data-driven credit risk assessment models that can be used to address the current and highly complex lending environments. The financial institutions are currently handling millions of loan applications that have various behavioural and financial characteristics, and common and most tree-based models are no longer sufficient to either capture nonlinear borrower patterns or cope with faces of harsh class imbalance. The result of these limitations is usually inaccurate classifications and poorer lending decisions. In order to mitigate these challenges, this research uses a DL model to predict credit default to improve its accuracy, fairness and scalability. In general, the work lead to more trustworthy automated underwriting systems and more powerful data-driven decision making in contemporary credit markets, and the advantages are detailed below:

- Extensive use of the Lending Club data, including all the pertinent demographics, financial, and loan-level data.
- A pre-processing pipeline, such as feature reduction, duplication deletion, outlier treatment, and missing-value imputation, categorical coding, and min-max normalization of the data to promote its quality and the model readiness.
- Use of the SMOTE to rectify the class disparity in an efficient manner and improve the model's capacity to identify minority-class (default) patterns.
- Implementation of an 80:20 stratified train-test partition to avoid sampling bias and to have reliable model testing.
- Architectural design and optimization of an ANN model that is trained on a large volume of tabular credit data, with which ResNet, FKNN, and AdaBoost classifiers are compared.
- F1-score, ROC, accuracy, precision, recall, and specificity, and AUC performance review on a big scale that shows that model's performance.
- Empirical results that deep-learning models have greater predictive stability and generalization to credit default prediction, which offer the practical use of tools to support more reliable risk assessment by lenders.

1.2. Justification and Novelty of the Paper

The necessity of the advanced predictive frameworks that are able to handle the size and complexity of contemporary credit datasets validates this paper's applicability. This study proposes an optimized ANN that predicts complex interactions between borrowers and lenders with the help of data analysis, unlike earlier studies that utilized ensemble machine learning and shallow neural networks. Its novelty is that it includes a full pipeline of pre-processing, uses SMOTE to balance classes and uses multi-metric performance analysis. Not only is the framework effective in increasing predictive performance, but also in improving reliability, which gives financial institutions an effective, practical and scalable solution in predicting loan default in the real-life credit risk management environment.

1.3. Structure of the Paper

The paper is structured as follows: Section II conducts a literature review of previous studies on ML and DL methods in credit risk assessment, Section III describes the methodology, which includes data collection, pre-processing, model execution, and evaluation processes, Section IV gives the results of the experiments, comparison and discussion of the model performance. Lastly, Section V is the conclusion of the study, and it presents suggestions for future research.

2. Literature Review

There are several ML and ensemble-based techniques to forecast credit risks, according to the literature study, limitations of DL adoption, imbalance treatment, and diversity of the datasets, and possibilities of improving prediction frameworks. Gasmi et al. (2025) A credit risk prediction model that assists bank employees in assessing client loan applications and categorizing them as "defaulter" or "non-defaulter" customers may be developed by looking at the intriguing decision variables. Based on the financial datasets from Taiwan and Australia, the multilayer perceptron (MLP) method was developed. Additionally, using the SMOTE approach, balance the data and improve the less-than-ideal outcomes. Use the R programming environment for statistical computations. With an average accuracy of 88.345% for the Taiwan dataset and 89.543% for the Australian dataset, the results demonstrate that economic, financial, demographic, and "behavioural indicators" have a considerable impact on credit risk and the effectiveness of the suggested model [14].

Li et al. (2025) To increase the precision of financial risk assessment models and offer practical information for risk control, forecast credit default risk using machine learning (ML). Because the current models lack the necessary accuracy, better ML algorithms and test models like XGBoost, LightGBM, RF, and DNN take their place. With an AUC of 0.784, the final model proves to be beneficial and lays the foundation for future developments in financial risk management [15]. Long (2024) The complexity of the global community and financial ecosystem makes it challenging to produce reliable credit risk prediction findings in the real application situation. To prevent the test dataset from becoming overfit, dropout was used, and the neural network module's composition and neuron count were established by experimentation. Ablation and stability tests show that the inter-dataset error can be controlled by the model to 0.021. The ideal number of neurons and hidden layers was shown in the ablation experiment. This approach outperformed previous classification techniques, as demonstrated by simulation testing, which indicated that its sensitivity and accuracy were 85.25% and 92.55%, respectively. In the last four years, tests were done based on real bank data [16]..

Aruleba and Sun (2024) purposes to improve classification performance by researching credit risk prediction using ensemble classifiers and the Synthetic Minority Over-sampling Edited Nearest Neighbour (SMOTE-ENN) technique. The ensemble classifiers include light gradient boosting machine (LightGBM), extreme gradient boosting (XGBoost), adaptive boosting (AdaBoost), and random forests. Class imbalance is addressed in the work by using Shapley additive explanations (SHAP) to understand model results and SMOTE-ENN, an ensemble classifier. The outcomes of the trial showed that the proposed strategy was better at categorization. Random Forest performed best on XGBoost has performed best on the German credit dataset has a recall of 0.930 and a specificity of 0.846, compared to the Australian dataset's 0.907 and 0.922 [17].

Jumaa, Saqib and Attar (2023) The article proposes that there is a novel approach to financial institution credit risk prediction utilizing ensemble ML models. After the pre-processing the data, then relevant characteristics are selected by using information gain technique to determine the relevance of the feature. The best 10 relevant characteristics are used to train the ML models. The proposed method for credit risk prediction is based on gradient boosting algorithms such as XGBoost, XGBoost RF, and CatBoost. The suggested method is compared to neural networks, Ad boost, Random Forest, or other advanced algorithms. Additionally, with maximum training accuracy of 93.7% and 93.6% and testing accuracy of 93.6% and 93.8%, the results show that gradient-boosting techniques like Xgboost and CatBoost are superior to other approaches [18].

Kanaparthi (2023) suggests a different method for predicting credit risk in financial organizations utilizing group ML models. Relevant characteristics are selected after the information is pre-processed, and their importance is determined based on the information gain method. 10 pertinent attributes are chosen to be utilized in the machine learning model training process. XGBoost, XGBoost RF, and CatBoost were among the gradient boosting algorithms that were included in the suggested approach to forecast credit risk. The suggested method is contrasted with many cutting-edge algorithms, including NN, RF, and Ad boost. Furthermore, with the highest testing accuracy of 93.6% and 93.8% and the best training accuracy of 93.7% and 93.6%, respectively, the results show that gradient-boosting algorithms like Xgboost and CatBoost outperform other methods [19].

The Table I have compared the major researches in the area of credit risk modelling, summarizing their contributions, methods, findings, limitations, and recommendations and presenting a systematic overview of the research that demonstrates gaps in research to prove that deep learning models are improved.

Table 1: Comparative Analysis of Prior Studies of Machine Learning-Based Credit Risk Modeling

Author(s)	Key Focus	Methods Used	Findings	Identified Limitations	Recommendations
Gasmi et al. (2025)	Credit risk prediction using demographic, financial, economic & behavioral variables	MLP neural network, SMOTE, R environment, Australian & Taiwan datasets	Accuracy: 89.543% (Australian), 88.345% (Taiwan). Features significantly influence credit risk.	Focused only on basic MLP; Limited datasets; No advanced deep learning; Basic imbalance handling; No interpretability	Explore advanced deep architectures (deep MLP, CNNs, hybrid DL); Use DL-native imbalance handling; Integrate explainability; Validate with more datasets
Li et al. (2025)	ML pipeline for credit default prediction	XGBoost, LightGBM, Random Forest, Deep NN, Ensemble	Final ensemble achieved AUC = 0.784	Deep learning not optimized; Tree-based models dominate; No efficiency evaluation	Develop optimized, efficient DL models; Improve feature engineering in DL context; Compare DL vs ML ensembles
Long (2024)	Deep learning + hierarchy analysis for credit risk	AHP + Deep Neural Networks; Dropout; Ablation testing	Accuracy = 92.55%; Sensitivity = 85.25%; Stable under stress tests; Error between datasets controlled to 0.021	AHP increases model complexity; Limited interpretability; Limited dataset diversity	Create lightweight DL models; Improve deep model interpretability; Test across global datasets
Aruleba & Sun (2024)	Ensemble classifiers + SMOTE-ENN + SHAP for interpretability in credit risk	RF, AdaBoost, XGBoost, LightGBM, SMOTE-ENN, SHAP	XGBoost achieved Recall = 0.930, Specificity = 0.846 on German dataset; RF achieved Recall = 0.907, Specificity = 0.922 on Australian dataset. SHAP enhanced	Focus primarily on tree-based ensembles; No deep learning models considered; Oversampling may distort minority class distribution	Explore deep learning models that handle imbalance inherently (focal loss/class weighting); Integrate explainable DL methods; Reduce over-dependence on oversampling

			feature transparency.		
Jumaa, Saqib & Attar (2023)	Predicting consumer loan default using deep learning	Keras-based Neural Network	Test accuracy = 95.2% (238/250 predictions correct)	Small, survey-derived dataset; Not generalizable; No comparison with other DL models	Use large real-world datasets; Compare multiple DL architectures; Apply strong regularization and cross-validation
Kanaparthi (2023)	Ensemble ML for credit risk	XGBoost, XGBoost RF, CatBoost, AdaBoost, Neural Networks	XGBoost achieved training accuracy = 93.7% and test accuracy = 93.6%; CatBoost slightly higher test accuracy (93.8%); XGBoost fastest	Neural networks underperformed; DL not deeply explored; Heavy reliance on boosting techniques	Investigate advanced deep learning architectures (DNN, CNN, TabNet); Improve DL training efficiency; Benchmark DL vs boosting models

3. Methodology

The methodology is aimed at creating a loan default prediction model based on DL, as seen in Figure 1. The method starts with the purchase of Lending Club data, then proceeds with a comprehensive data preparation process to ensure the data's consistency and quality. This involves data cleaning of missing values, coding categorical data, outliers, feature scaling and the class imbalance to avoid biased prediction. To ensure exceptional performance on unknown data, the collected data is then divided into training and testing sets, with 80% going to the former and 20% to the latter. A training set is run in an ANN to acquire the complex patterns of loan defaults. Evaluation of model performance is done based on various measures, such as F1-score, which has provided a thorough examination of predictive performance, accuracy, precision, recall, and specificity. This process of work guarantees quality forecasts to make wise choices in credit risk management.

3.1. Data Collection

Lending Club Loan Dataset is a publicly accessible dataset that includes 2,925,492 rows and 142 features allowing One of the most thorough methods for analyzing credit risk. The rows are a single loan issued in 2007-2020 and the 142 features include a broad selection of borrower, financial and loan level features. These consist of the loan amount, interest rate, length of work, yearly income, loan status information, credit history indicators, and the debt-to-income ratio. The dataset is useful in in-depth modelling of loan performance and has been extensively used in research on ML, financial risk prediction and optimization of lending decisions.

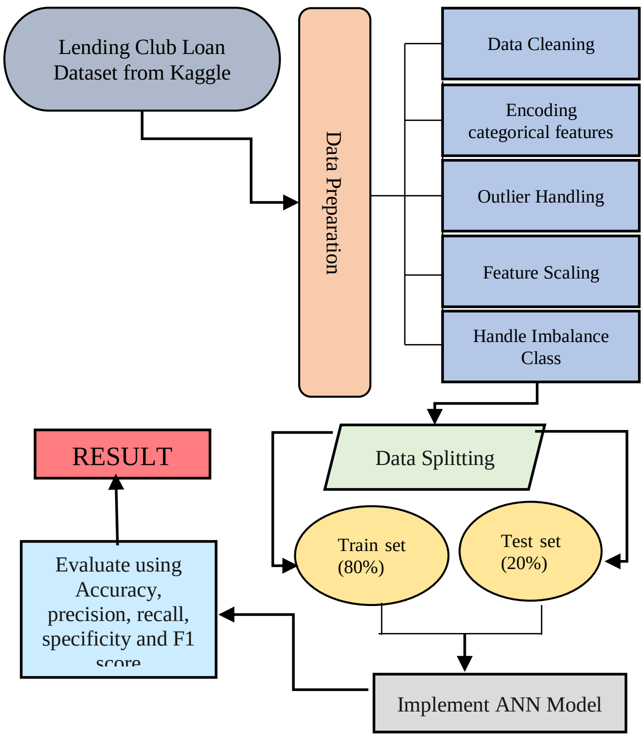

Fig 1: Workflow for Deep Learning-Based Loan Default Prediction

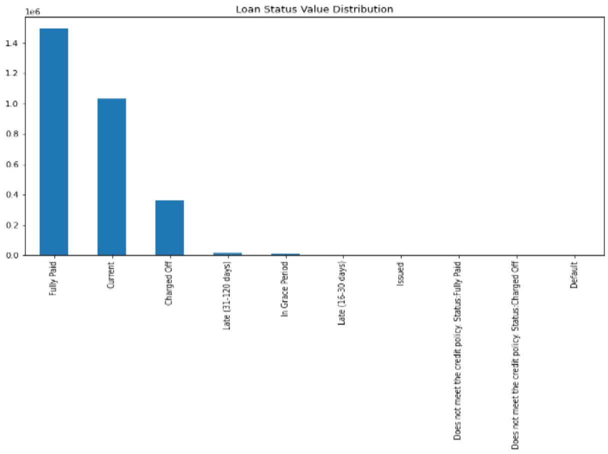

Fig 2: Loan Status Value Distribution

Figure 2 provides a summary of the amount of Lending Club loans by their current status. The most important categories are Fully paid and Current with Fully paid having the largest volume of a total of more than 1.4 million and Current right after it with a total of more than 1.0 million. The third-largest category, about 0.35 million, are the loans that were charged off. Other statuses such as Late, In Grace Period, Issued and policy related ones are considerably smaller. This implies that the huge proportion of loans in this dataset have been comfortably repaid or in good standing. In general, this distribution points to the recurring success and the active size of portfolio of the lending business of Lending Club.

The correlation matrix of different variables in the Lending Club dataset between the years 2007 and Q3 2020 is shown in Figure 3. The level of colours and number in the matrix indicates the Pearson correlation coefficients between two variables. There are the most positive correlations (which are more similar to +1 and are therefore bright yellow) along the diagonal, since each variable is completely correlated with itself. Additionally, there is a strong positive association between a few loan attributes; that is, the first two variables mentioned above (probably concerning the loan amount or interest rate) have a strong positive correlation of 0.95. The others feature pairs exist in moderate-weak relationships (that is, they are closer to 0, and hence are greenish-blue), which support the fact that most of the features are comparatively independent.

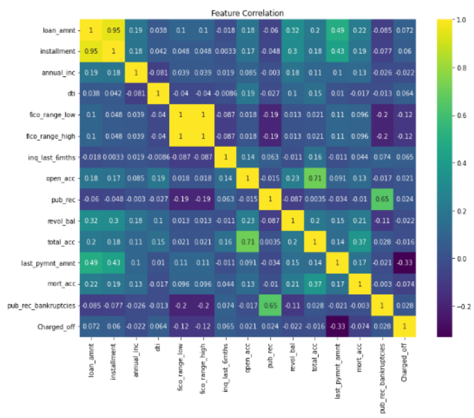

Fig 3: Correlation Heatmap of Data Features

3.2. Data Cleaning

The dataset was first of all cleaned and reduced to guarantee the data quality and reliability before implementing any modelling techniques. Controversies with too many missing values were eliminated, and the remaining missing entries were treated systematically. The following actions were taken to improve the models' performance:

- Eliminate sparse features: Dropping all the features that have over 50% missing values reduce the size of the dataset by 142 to 107 features.
- Impute numerical missing values: Utilizing the median, complete the numerical feature's columns with missing values.
- Impute categorical missing values: Replace categorical missing values with the mode of the column or with a new category called unknown.
- Check the consistency of data: All imputed values should be properly absorbed and none of the absent entries should be left.
- Document dropped and imputed features: Maintain log of all the eliminated columns and imputed for the reproducibility.

3.3. Encoding Categorical Features

All the categorical features were converted into numeric values in order to feed the dataset into the artificial neural network. One-hot encoding was used to encode the low-cardinality categorical variables, and each category was turned into a binary column, and the target variable, loan-status, was coded into the numerical labels (label encoding). The reason of this step is to make sure that the model is capable of handling categorical data without creating unwanted ordinal associations that compromise the data integrity of the data to be used in correct training and prediction.

3.4. Feature Scaling

Certain models may be impacted by features in a dataset with a varied range. Using scaling methods, the features were scaled to address this issue. Data is scaled to the interval [0, 1] [20]. It can be affected by outliers even though it is not as sensitive to them as the regular scaler. Here's how it's calculated: as Equation (1)

$$n = \frac{n - minimum(n)}{maximum(n) - minimum(n)} \quad (!)$$

3.5. Handling Class Imbalance

The imbalance in the target variable loan status was observed to be significant between the two major categories that are Fully Paid and Charged Off. To overcome this, a synthetic Minority Oversampling Technique (SMOTE) was used on the dataset, which synthetically produced samples of the underrepresented class of Charged Off. This balanced the data, and the model was able to learn effectively using both classes. Figure 4 demonstrates the distribution of classes before and after the application of SMOTE where it can be seen that minority classes became more representative in terms of numbers and the distribution became more balanced in terms of representation on the model training.

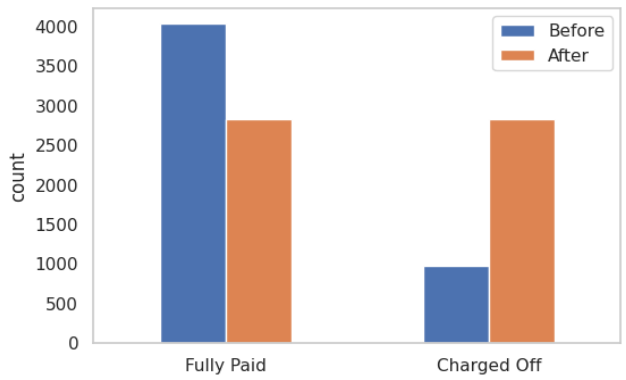

Fig 4: Loan Status Counts Before and After SMOTE

3.6. Data Splitting

The dataset was then partitioned into 80:20 training and testing subsets after all pre-processing steps were complete, in order to maintain the desired class distribution in both training and testing data, stratified sampling was used.

3.7. Implement the Artificial Neural Network Model

An ANN is a nonlinear technique that offers a fresh take on linear approaches, particularly when the dataset has intricate correlations between the nonlinear variables' independence. To model a nonlinear connection between inputs and outputs, one kind of learning system is an artificial neural network (ANN) [21]. Because it is challenging to obtain data from an internal system, they are often known as "black box" systems [22]. ANN are ML systems that mimic the composition and functionality of organic neurons. In the same manner that a neuron alters its states to carry out a cognitive activity, artificial neural networks, or ANNs, accomplish a task by modifying its parameters. A collection of neurons arranged according to a predetermined topology makes up a network. Weights, which are functions that depict the activity of the neural network, are linked to neurons and control the strength of information flow.

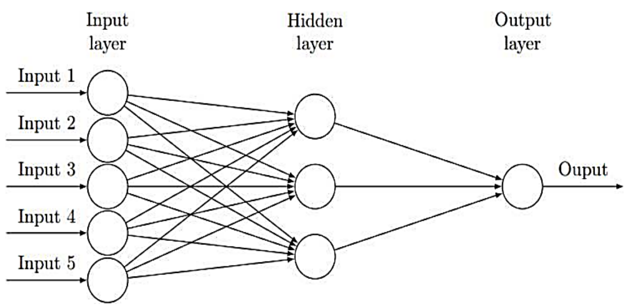

Fig 5: Basic Structure of Artificial Neural Network

Figure 5 three layers make up an ANN, the input layer, which shows neurons receiving input impulses, the output layer, and the hidden layer. The data is then moved to the concealed layer, which is the next tier of the layer. Depending on the extent of the connections between neurons, information is weighed before moving on to the next layer. Every neuron has a transfer function and a minimal value (threshold value) that causes the neuron to fire. A hidden layer, which can have many layers, generates output neurons by adding up the weights by the sum of the input neurons [23]. The method of generating output involves two steps: Before all values are combined, each input is multiplied by the weight on the relevant link; the activation function is then applied to all inputs.

$$y_i = \sum_{j \in I} W_{j,i} a_j \tag{2}$$

$$a_i = g(y) \tag{2}$$

As (2) denotes input evaluation and (3) denotes activation function, $W_{j,i}$ stands for weights on the link between j and , i, and a_j is the neuron j s activation function.

3.8. Evaluation Metrics

To assess unbalanced datasets, accuracy alone is insufficient. As a result, assess the effectiveness of ML techniques using a different assessment metric known as the confusion matrix. As seen in Figure 6, the confusion matrix is a particular table arrangement that may display the classifier result.

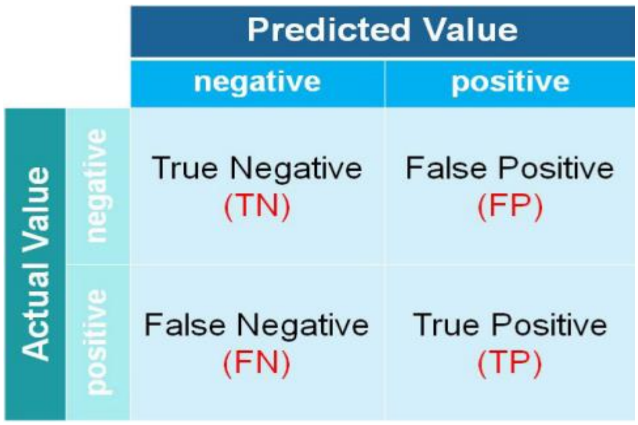

Fig 6: Confusion Matrix[24]

TP (true positive) and TN (true negative) are examples of classifiers that predict accuracy. In the meanwhile, Type I and Type II errors are represented by examples of classifiers that anticipate incorrectness, which are known as FP (false positive) and FN (false negative). The metrics are given below with their formulas.

3.8.1. Accuracy

The simplest and most widely used statistic for classification issues is accuracy. It is determined by dividing the number of forecasts by the number of accurate projections. When talking about accuracy, it is also essential to take into account the true negative rate (TNR), true positive rate (TPR), false negative rate (FNR), and false positive rate (FPR). Calculating accuracy may be done using Equation (4):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

3.8.2. Precision

Precision makes an effort to respond to the query, "What percentage of positive identifications were actually accurate?" Prediction is the foundation of precision [25]. The following is the mathematical Equation (5) for precision:

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

3.8.3. Recall

Recall makes an effort to respond to the query, "What proportion of TP were detected correctly?" The truth serves as the basis for the recall. The recall is represented by the following mathematical Equation (6):

$$Recall = \frac{\text{TP}}{TP+FN} \tag{6}$$

3.8..4. Specificity

Specificity is a measure of binary classification (i.e., medical testing) that quantifies the accuracy of a model's negative case predictions [26]. That is, it informs us of the percentage of true negative cases (correctly identified cases of non-disease) in all real negative cases. It is also referred to as the TN rate. The following is the mathematical Equation (7) for precision:

$$Specificity = \frac{TN}{TN+FP} \tag{7}$$

3.8.5. F1 Score

The F1 score is a harmonic average of recall and goes from 0 to 1 and accuracy, is another performance statistic. The model performs better overall if its F1 score is greater, where 0 represents the lowest performance, and 1 represents the best. It only reaches its optimal level of 1 when recall and accuracy are both 100%. If 1 of these is equal to 0, the F1 score has its lowest value of 0. Equation (8) shows the calculation:

$$F1 = 2 \times \left(\frac{(Precision \times Recall)}{(Precision + Recall)}\right) \tag{8}$$

3.8.6. AUC

An essential assessment metric that gauges a classification model's overall effectiveness is the receiver operating characteristic curve's area under the curve, or AUC. The area under the ROC curve shows the trade-off between the TPR and FPR of the categorization threshold values. Plotting FPR on the x-axis and TPR on the y-axis, the ROC curve shows how well the model distinguishes between positive (default) and negative (non-default) classes.

These measures are used in credit risk management to evaluate how well the model performs and forecasts.

4. Results & Discussion

The section is a report of experimental results of the DL model developed to predict default loans in credit management. This investigation is based on the findings of an ANN model applied to the data of the Lending Club. Each experiment was performed on a computer platform with Windows 10 and the 3.3 GHz Intel dual-core i6 CPU with 1 TB of RAM. The model was assessed using the following metrics: F1-score, as shown in Table II, sensitivity (recall); specificity; and accuracy and precision. The measurements assess the effectiveness of the approach by comparing defaulting and non-defaulting debtors. The ANN model attained a 98.92% accuracy with a recall of 99.2% showing that it strongly identified cases of a default. A specificity of 99.1% and a precision of 98.08% indicates that there are low false predictions and it is a reliable indication of credit risk since the 98.63% F1-score indicates a robust balance between memory and accuracy.

Table 2: Performance of the Ann Model for Default Loan Prediction Using the Lending Club Dataset

Metric	ANN model
Accuracy	98.92
Precision	98.08
Sensitivity (Recall)	99.2
Specificity	99.1
F1-score	98.63

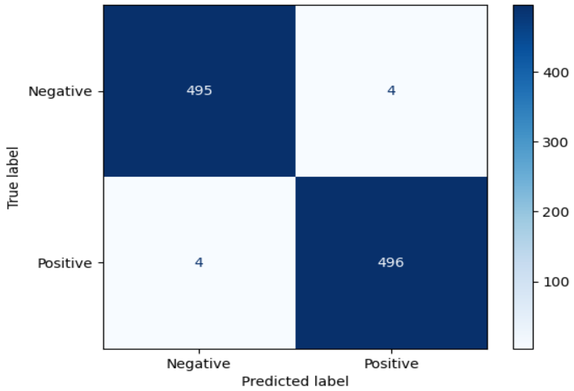

Fig 7: Confusion Matrix of the ANN Model for Loan Default Prediction

The confusion matrix of ANN model in credit risk management has high predictive power in Figure 7. The model was found to be highly reliable in the discrimination between the two categories of loan cases with 495 cases of non-defaults (TN) and 496 cases of default (TP) out of all cases of loans. There were 4 times fewer non-defaults wrongly identified as defaults (FP) and 4 wrongly identified as non-defaults as defaults (FN). These slight misclassifications indicate the general accuracy and effectiveness of the model. These high and balanced rates at true prediction points indicate that the ANN model is appropriate when it comes to credit risk, which would be reliable in making decisions when it comes to loan approvals.

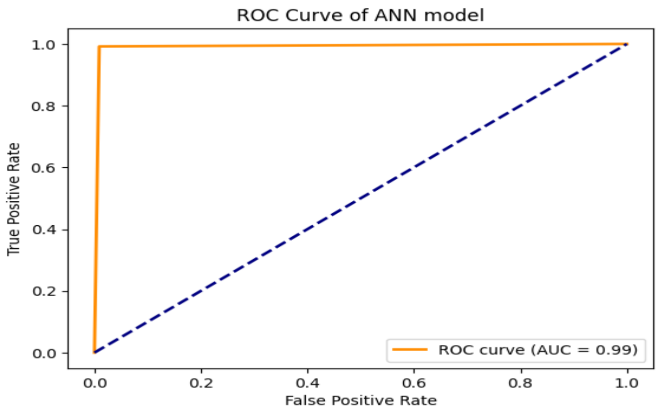

Fig 8: ROC Curve of ANN Model for Loan Default Prediction

The ROC curve of the ANN model in credit risk management is presented in Figure 8, and this clearly shows how the model demonstrates its superb classification ability. The curve used to plot the FPR as a function of the threshold value and the TPR (Recall). The orange line is directly following the upper-left section of the chart, which shows that there is almost perfect discrimination among the classes. Above all, the model has an AUC of 0.99. As the AUC is between 0 and 1 with 1 being the perfect performance, AUC of 0.99 proves that the ANN model is highly effective to differentiate between default and non-default borrowers and proves its outstanding applicability to credit risk assessment.

4.1. Comparative Analysis and Discussion

This section compares various ML models in order to identify the best and most efficient model for credit risk management's loan default prediction. A comparative analysis of four models, namely ResNet, FKNN, AdaBoost and ANN, are employed to predict loan default in table III ANN model performs better than other models, having the highest accuracy (98.92%), precision (98.08%), recall (99.2%), and F1-score (98.63%), which demonstrates the best predictive ability. AdaBoost shows good performance with a little bit low scores compared to ANN, FKNN and ResNet have moderate results. All in all, the table underscores the strength and efficacy of the ANN to accurately detect a defaulting and a non-defaulter borrower.

Table 3: Performance Comparison of Machine Learning Models for Loan Default Prediction

Models	Accuracy	Precision	Recall	F1-score
ResNet[27]	76.45	80.65	91.88	85.90
FKNN[28]	87.14	89.18	88.05	88.61
AdaBoost[29]	91.30	86.36	95.0	90.48
ANN	98.92	98.08	99.2	98.63

The analysis of the experiment reveals that ANN model is evidently more effective than ResNet, FKNN and AdaBoost in classifying loan defaults and that it has higher general classification performance. The confusion matrix as well as ROC curve also suggest that the model is reliable and strong, with significantly less misclassifications. These findings support the ability of the ANN to give nonlinear complex trends of huge financial information. Therefore, deep learning is a highly effective procedure in the measurement of credit risk, which provides lenders with more effective information to reduce the risk of default and improve their performance in decisions.

5. Conclusion & Future Work

The credit default prediction methods of deep-learning have clear performance advantages compared to traditional and ensemble methods. The ANN model scored much higher on all the evaluation measures of accuracy of 98.92, precision of 98.08, recall of 99.2, specificity of 99.1 and F1-score of 98.63, and an AUC of 0.99 which is a good figure of discriminatory capability. The capability of the model to provide such results demonstrates how it can not only extrapolate complicated non-linear relationships between borrowers, but also reduce poor borrower classification, both of defaulting and non- defaulting applicants. Properly credited with its vast preprocessing and practice of handling the class imbalance, the ANN is a very reliable and useful tool of alleviating lending decisions, lowering exposure to default risk, and enhancing the efficacy of the credit risk as a unit. These findings can be developed further in future practice. Additional tools for interpretability, such explainable AI tools like SHAP or LIME, would help to make predictions more transparent and can help institutions to understand some of the key factors affecting predictions. It may be improved by generalizing and expanding on the predictive depth with more data sources such as credit bureau histories or behavioral variables or macroeconomic indices. More can be more conscious of the complex interaction and interdependence among borrowers by investigating more advanced architecture like Transformers, attention, or hybrid DL ensembles. Finally, the model assessment based on multi-year, real-time, or cross-institution data would provide even more extrinsic validation and be used more efficiently in a broader variety of financial contexts.

References

[1] N. Malali, “Exploring Artificial Intelligence Models for Early Warning Systems with Systemic Risk Analysis in Finance,” in *2025 International Conference on Advanced Computing Technologies (ICoACT)*, IEEE, Mar. 2025, pp. 1–6. doi: 10.1109/ICoACT63339.2025.11005357.

[2] D. Patel, “Enhancing Banking Security: A Blockchain and Machine Learning- Based Fraud Prevention Model,” *Int. J. Curr. Eng. Technol.*, vol. 13, no. 06, pp. 576–583, 2023, doi: 10.14741/ijcet/v.13.6.10.

[3] R. Q. Majumder, “A Review of Anomaly Identification in Finance Frauds Using Machine Learning Systems,” *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 5, no. 10, pp. 101–110, Apr. 2025, doi: 10.48175/IJARSCT-25619.

[4] O. Adekunle, S. Y. Alalade, T. Agbatogun, and C. Abimbola, “Credit risk management and financial performance of selected commercial banks in Nigeria,” *J. Econ. \& Financ. Stud.*, vol. 3, no. 01, pp. 1–9, 2015.

[5] S. J. Wawge, “A Survey on the Identification of Credit Card Fraud Using Machine Learning with Precision, Performance, and Challenges,” *Int. J. Innov. Sci. Res. Technol.*, vol. 10, no. 4, pp. 3345–3352, May 2025, doi: 10.38124/ijisrt/25apr1813.

[6] B. R. Ande, “Federated Learning and Explainable AI for Decentralized Fraud Detection in Financial Systems,” *J. Inf. Syst. Eng. Manag.*, vol. 10, no. 35s, pp. 48–56, 2025, doi: 10.52783/jisem.v10i35s.5921.

[7] R. Ma, “German Credit Risk Prediction Using Machine Learning Models,” in *2025 3rd International Conference on Image, Algorithms, and Artificial Intelligence (ICIAAI 2025)*, 2025, pp. 283–292. doi: 10.2991/978-94-6463-823-3_27.

[8] H. Kali, “Optimizing Credit Card Fraud Transactions identification and classification in banking industry Using Machine Learning Algorithms,” *Int. J. Recent Technol. Sci. Manag.*, vol. 9, no. 11, pp. 85–96, 2024.

[9] S. B. Shah, “Advanced Framework for Loan Approval Predictions Using Artificial Intelligence-Powered Financial Inclusion Models,” in *2025 IEEE Integrated*

STEM Education Conference (ISEC), 2025, pp. 1–10. doi: 10.1109/ISEC64801.2025.11147327.

[10] R. Dattangire, R. Vaidya, D. Biradar, and A. Joon, "Exploring the Tangible Impact of Artificial Intelligence and Machine Learning: Bridging the Gap between Hype and Reality," in *2024 1st International Conference on Advanced Computing and Emerging Technologies (ACET)*, IEEE, 2024, pp. 1–6. doi: 10.1109/ACET61898.2024.10730334.

[11] N. Prajapati, "The Role of Machine Learning in Big Data Analytics: Tools, Techniques, and Applications," *ESP J. Eng. Technol. Adv.*, vol. 5, no. 2, 2025, doi: 10.56472/25832646/JETA-V5I2P103.

[12] M. Ntow-Gyamfi and S. S. Boateng, "Credit risk and loan default among Ghanaian banks: An exploratory study," *Manag. Sci. Lett.*, vol. 3, no. 3, pp. 753–762, Mar. 2013, doi: 10.5267/j.msl.2013.01.015.

[13] V. Verma, "Deep Learning-Based Fraud Detection in Financial Transactions : A Case Study Using Real-Time Data Streams," vol. 3, no. 4, pp. 149–157, 2023, doi: 10.56472/25832646/JETA-V3I8P117.

[14] I. Gasmi, S. Neji, N. Mansouri, and M. Soui, "Bank credit risk prediction using machine learning model," *Neural Comput. Appl.*, vol. 37, no. 16, pp. 10333–10350, Jun. 2025, doi: 10.1007/s00521-025-11044-5.

[15] W. Li, L. Yao, W. Li, and C. Ji, "Visualizing Credit Default Risk: Insights Through Machine Learning," in *2025 7th International Conference on Software Engineering and Computer Science (CSECS)*, IEEE, Mar. 2025, pp. 1–6. doi: 10.1109/CSECS64665.2025.11009748.

[16] X. Long, "Development and Implementation of a Multilayer Deep Learning-Based Bank Credit Risk Forecasting System," *J. Inf. \& Knowl. Manag.*, vol. 23, no. 04, p. 2450051, 2024.

[17] I. Aruleba and Y. Sun, "Effective Credit Risk Prediction Using Ensemble Classifiers With Model Explanation," *IEEE Access*, vol. 12, pp. 115015–115025, 2024, doi: 10.1109/ACCESS.2024.3445308.

[18] M. Jumaa, M. Saqib, and A. Attar, "Improving Credit Risk Assessment through Deep Learning-based Consumer Loan Default Prediction Model," *Int. J. Financ. Bank. Stud.*, vol. 12, no. 1, pp. 85–92, Jun. 2023, doi: 10.20525/ijfbs.v12i1.2579.

[19] V. Kanaparthi, "Credit Risk Prediction using Ensemble Machine Learning Algorithms," in *2023 International Conference on Inventive Computation Technologies (ICICT)*, IEEE, Apr. 2023, pp. 41–47. doi: 10.1109/ICICT57646.2023.10134486.

[20] A. Akinjole, O. Shobayo, J. Popoola, O. Okoyeigbo, and B. Ogunleye, "Ensemble-Based Machine Learning Algorithm for Loan Default Risk Prediction," *Mathematics*, vol. 12, no. 21, p. 3423, Oct. 2024, doi: 10.3390/math12213423.

[21] P. Sharifi, V. Jain, M. Arab Poshtkohi, E. Seyyedi, and V. Aghapour, "Banks credit risk prediction with optimized ANN based on improved owl search algorithm," *Math. Probl. Eng.*, vol. 2021, no. 1, p. 8458501, 2021.

[22] D. K. Gupta and S. Goyal, "Credit Risk Prediction Using Artificial Neural Network Algorithm," *Int. J. Mod. Educ. Comput. Sci.*, vol. 10, no. 5, pp. 9–16, May 2018, doi: 10.5815/ijmecs.2018.05.02.

[23] V. Boguslauskas and R. Mileris, "Estimation of credit risk by artificial neural networks models," *Eng. Econ.*, vol. 64, no. 4, 2009.

[24] Y.-R. Chen, J.-S. Leu, S.-A. Huang, J.-T. Wang, and J.-I. Takada, "Predicting Default Risk on Peer-to-Peer Lending Imbalanced Datasets," *IEEE Access*, vol. 9, pp. 73103–73109, 2021, doi: 10.1109/ACCESS.2021.3079701.

[25] V. Chang, S. Sivakulasingam, H. Wang, S. T. Wong, M. A. Ganatra, and J. Luo, "Credit Risk Prediction Using Machine Learning and Deep Learning: A Study on Credit Card Customers," *Risks*, vol. 12, no. 11, p. 174, Nov. 2024, doi: 10.3390/risks12110174.

[26] A. Khashman, "Credit risk evaluation using neural networks: Emotional versus conventional models," *Appl. Soft Comput.*, vol. 11, no. 8, pp. 5477–5484, 2011.

[27] J.-Y. Kim and S.-B. Cho, "Towards Repayment Prediction in Peer-to-Peer Social Lending Using Deep Learning," *Mathematics*, vol. 7, no. 11, 2019, doi: 10.3390/math7111041.

[28] Z. Hassani, M. Alambardar Meybodi, and V. Hajihashemi, "Credit Risk Assessment Using Learning Algorithms for Feature Selection," *Fuzzy Inf. Eng.*, vol. 12, no. 4, pp. 529–544, 2020, doi: 10.1080/16168658.2021.1925021.

[29] I. R. Berrada, F. Barramou, and O. B. Alami, "Towards a Machine Learning-based Model for Corporate Loan Default Prediction," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 3, pp. 565–573, 2024, doi: 10.14569/IJACSA.2024.0150357.