



## **Enhanced Credit Card Default Prediction Using a Multilayer Perceptron With Z-Score–Based Outlier Handling on The UCI Dataset**

R. Yakubu<sup>1</sup>, A.A Abubakar<sup>2</sup>, D.G. Lazarus<sup>1</sup>, A.A. Babajo<sup>1</sup>

### **Abstract**

Credit card default prediction remains a central challenge for financial institutions, particularly due to data irregularities, class imbalance, and outlier-driven distortions. While Deep Neural Networks (DNNs) have been widely adopted, their performance is often sensitive to noisy data and computationally demanding. This study investigates whether integrating z-score–based outlier removal with a Multilayer Perceptron (MLP) can enhance predictive performance on the UCI Default of Credit Card Clients dataset. The methodology involved removing outliers using the z-score method per class, standardizing the dataset, partitioning the data into training and test sets, and developing a two-hidden-layer MLP model. After detecting and removing outliers separately within each class, a total of 3,785 anomalous observations were eliminated, resulting in 26,215 usable samples. The cleaned dataset was standardized, split using a stratified 70:30 ratio, and used to train an MLP with two hidden layers. Model performance was evaluated using accuracy, precision, recall, F1-score, specificity, G-mean, and AUC. The MLP achieved an accuracy of 83%, precision of 0.78, recall of 0.67, and F1-score of 0.70, strong specificity (0.96), confirming its ability to correctly classify non-defaulters. AUC score of 0.79 which indicates good overall discriminative power. The confusion matrix showed that while most non-defaulters were correctly identified, a notable number of defaulters were misclassified as non-defaulters. Compared to the DNN, the MLP outperformed across all metrics, with the most substantial improvements observed in recall (0.67 vs. 0.39) and F1-score (0.70 vs. 0.48). These findings demonstrate that a relatively simple MLP architecture, when combined with robust outlier handling, can outperform more complex models such as DNNs.

---

✉ R. Yakubu: [rukayyayakubumar5@gmail.com](mailto:rukayyayakubumar5@gmail.com)

<sup>1</sup>Department of Informatics, Kaduna State University, Nigeria

<sup>2</sup>Department of Secure Computing, Kaduna State University, Nigeria

---

**Keywords:** Machine Learning; Deep Learning; Credit Default; Z-Score handling; Deep Neural Network

### Article history

Received: October 29, 2025

Received in revised form: December 19, 2025

Accepted for publication: December 21, 2025

Available online: December 23, 2025

## 1.0 Introduction

Credit card default prediction plays a critical role in maintaining financial stability, preventing loan losses, and improving decision-making processes in lending institutions. As credit usage continues to rise globally, banks face increasing exposure to repayment uncertainty. Traditionally, credit risk assessment relied on linear statistical models; however, the complexity and non-linearity inherent in financial behaviour have encouraged a shift toward machine learning and deep learning techniques (Bello, 2020; Shu & Ye, 2022; Chouksey et al., 2023).

Previous studies have applied a wide range of machine learning (ML) and deep learning (DL) models to credit risk and related financial prediction problems. Mestirri (2024) evaluated six models such as Linear Discriminant Analysis, Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and Deep Neural Network (DNN), using 688 samples. The DNN achieved the highest accuracy (83.6%), demonstrating the advantage of nonlinear models, although the small dataset and limited preprocessing constrained generalizability. Similarly, Alarfaj et al. (2022) compared classical ML models with Convolutional Neural Networks (CNNs) for fraud detection. While CNNs performed best, achieving 99.9% accuracy, 85.71% F1-score, 93% precision, and 98% AUC, the study did not evaluate simpler architectures such as Multilayer Perceptions (MLPs) nor address outlier handling. In a related study, Bhandary and Ghosh (2025) compared a DNN with several classical and ensemble models on the UCI Credit Card Default dataset, reporting that the DNN achieved the highest performance with an accuracy of 81.8%.

Although DNNs have shown strong performance in complex prediction tasks, they face limitations such as high computational cost, interpretability challenges, and sensitivity to noisy or irregular financial data (Bhandary & Ghosh, 2025; Talaat et al., 2023). One particular challenge in credit datasets is the presence of outliers, observations that deviate significantly from the majority of data points, which may result from reporting errors, rare behavioural patterns, or extreme financial actions. These outliers can distort learning processes and cause models to misclassify minority class samples, particularly defaulters (Chang et al., 2024; Nkambule et al., 2024; Dube & Verster, 2024).

Despite extensive use of DNNs and other advanced models, few studies have systematically examined whether robust preprocessing, particularly Z-score-based outlier removal, a statistical method that identifies observations beyond a set number of standard deviations from the mean can enable simpler models such as MLPs to outperform deeper architectures. This gap is significant because simpler models offer lower computational cost, easier deployment, and improved interpretability.

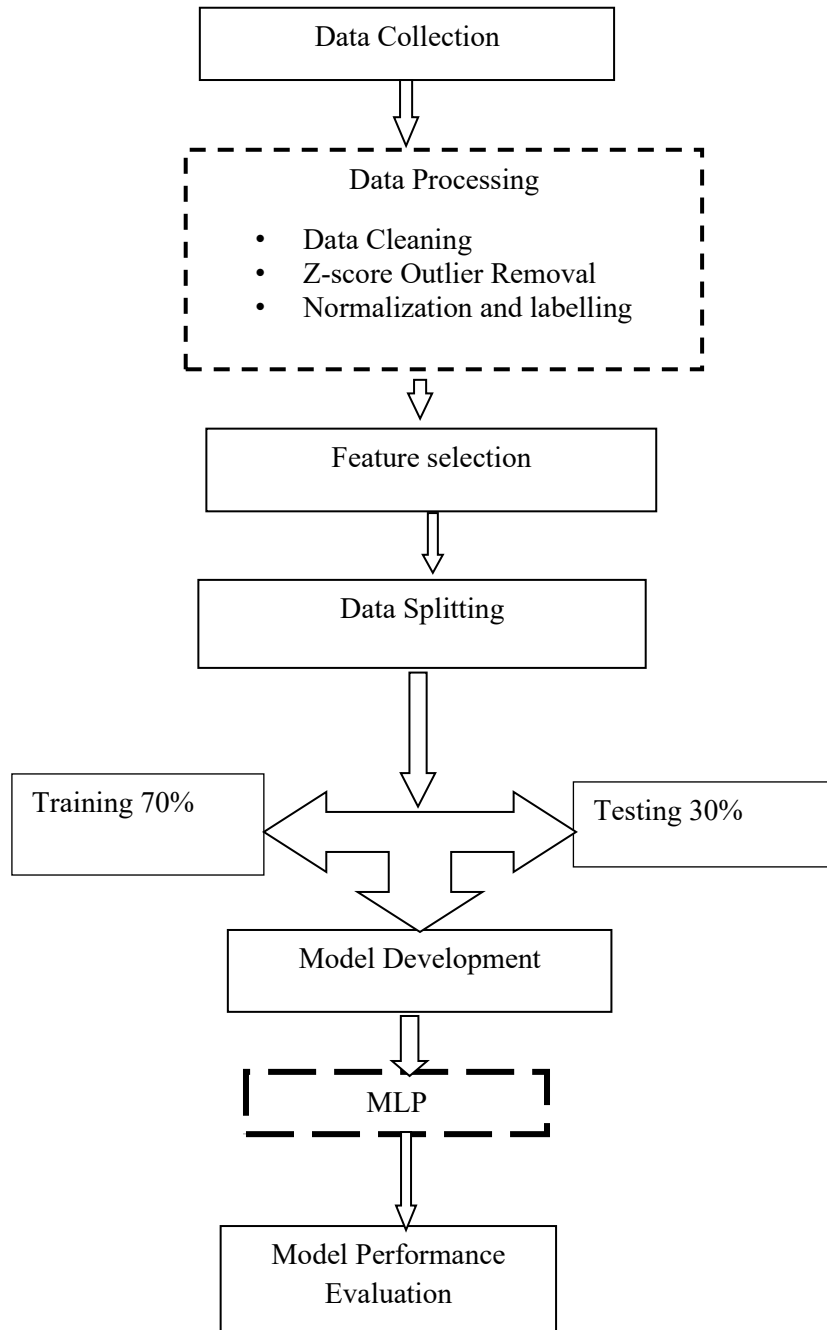
This study investigates whether integrating Z-score-based outlier removal with an MLP, a feedforward neural network with one or more hidden layers improves classification performance on the UCI Credit Card Default dataset. The model is evaluated against a benchmark DNN (Bhandary & Ghosh, 2025), with the goal of demonstrating that data quality improvements can substantially influence model performance, potentially more than increasing model complexity.

## 2.0 Materials and Methods

### 2.1 Research Design

This study adapted an experimental research design (Bhandary & Ghosh, 2025). In order to evaluate the performance of a Multilayer Perceptron model in predicting credit card default and evaluating its performance on a

publicly available dataset. The design was guided by the need to systematically preprocess the dataset, implement the proposed neural network model, and evaluate its predictive performance against results from prior studies (Bhandary & Ghosh, 2025). The research process shown in Figure 1.



**Fig. 1:** Research Process

### 2.2 Dataset Description

The dataset used is the Taiwan credit card default dataset, which consists of 30,000 observations and 24 predictor variables,

including demographic information, payment history, bill statements, and repayment amounts, including demographic attributes, payment history, bill amounts, and default status. The

target variable is binary, indicating whether a client defaulted on credit card payments in the subsequent month (1 = default, 0 = non-default).

## 2.3 Data Preprocessing

### 2.3.1 Outlier Removal using Z-score

Outliers can adversely affect the performance of machine learning models, particularly in financial datasets where anomalous transactions or reporting errors may occur (Lartey et al., 2024; Han & Joe, 2025; Blazek & Duricova, 2025). To address this, outliers were detected using the standard z-score method:

$$Z = \frac{x - \mu}{\sigma} \quad (1)$$

where  $x$  is the individual data point or observation being evaluated,  $\mu$  is the mean of all data points in the dataset and represents the central tendency of the data,  $\sigma$  is the standard deviation of the dataset, which measures the dispersion of the data points around the mean and  $Z$  is the standardized score (Z-score) obtained by normalizing  $x$  with respect to  $\mu$  and  $\sigma$ .  $Z$  expresses the relative position of an observation in terms of standard deviation units. Observations with feature values exceeding three standard deviations from the mean ( $|Z| > 3$ ) were marked as outliers and excluded from further analysis. Outlier detection was applied separately to each class (default and non-default), preventing distortion of class distribution. The procedure ensured that class balance was preserved while improving data quality for model training. The implementation removed a total of 3,785 outliers, resulting in a total of 26, 215 usable samples for modelling. This method was chosen because Z-score is widely used for continuous financial attributes, Class-specific removal avoids eliminating rare but valid defaulter patterns.  $|Z| > 3$  aligns with statistical convention for extreme deviations.

### 2.3.2 Feature Normalization

Following outlier removal, the predictor variables were standardized using the Standard Scaler from the Scikit-learn library.

Standardization transformed the features to have zero mean and unit variance, thereby ensuring that all variables contributed equally to model learning and preventing bias from features with larger scales.

This step is critical because features in financial datasets often have different scales (e.g., bill amounts in thousands vs. age in years). Without normalization, features with larger magnitudes could disproportionately influence the MLP's weight updates, leading to biased learning. Standardization ensures that all features contribute equally, improves the convergence speed of the model during training, and prevents vanishing or exploding gradients in neural networks.

### 2.3.2 Train-Test Split

The processed dataset was partitioned into training and test sets using a 70:30 ratio, with stratification to preserve the proportion of default and non-default cases. The training set was used for model fitting, while the test set served as unseen data for model evaluation.

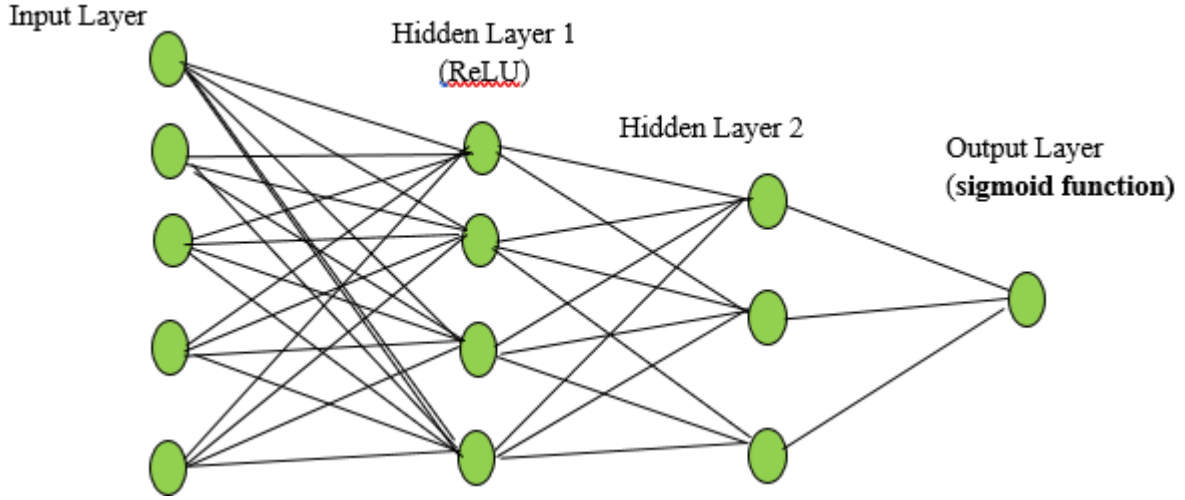
### 2.3.3 MLP Model Development

The MLP model was constructed using the TensorFlow / Keras framework in Python 3 on google Collaboratory platform, which provides efficient and accessible cloud-based environment, Pandas, NumPy was used for the data preprocessing, Stats for z-score outlier removal, geometric mean score to handle class imbalance and Matplotlib for visualization, and confusion matrix.

The architecture comprised an input layer corresponding to the number of predictor variables, two hidden layers with 32 and 16 neurons respectively, and an output layer with a single neuron using the sigmoid activation function for binary classification. Rectified Linear Unit (ReLU) activation functions were employed in the hidden layers due to their efficiency in mitigating vanishing gradient problems.

The model was compiled using the RMSprop optimizer with a learning rate of 0.0001, rho of 0.8, and momentum of 0.1. The loss function selected was binary cross-entropy, which is well-suited for binary classification problems. Seed for model initialization: 42

Accuracy was included as a primary metric during training. Figure 2 shows the current study's Proposed MLP Model Architecture for Credit Card Default Prediction adopted from Bhandary and Ghosh (2025).



**Fig 2:** Proposed MLP Model Architecture for Credit Card Default Prediction (Adopted from Bhandary & Ghosh, 2025).

## 2.4 Model Training

The model was trained on the processed training dataset for 60 epochs with a batch size of 32. A validation split of 20% of the training data was used to monitor generalization performance during training. The training process involved iterative adjustments of network weights through backpropagation to minimize the loss function.

### 2.4.1 Visualization of Training Process

To assess convergence behavior, training and validation accuracy and loss were plotted across epochs. These visualizations provided insights into the model's learning dynamics and helped detect issues such as overfitting or underfitting.

## 2.5 Model Performance Evaluation

The trained MLP model was evaluated on the test dataset using multiple performance metrics to provide a comprehensive assessment. These metrics include accuracy, precision, recall, F1-score, G-mean score, and AUC.

**Accuracy:** Measures the overall proportion of correct predictions out of all predictions, serving

as a primary indicator of the model's effectiveness.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

**Precision:** Highlights the model's ability to avoid false positives, by calculating the ratio of true positive predictions to the total predicted positives for each class.

$$Precision = \frac{TP}{TP + FN} \quad (3)$$

**Sensitivity/Recall:** Computes the ratio of true positive predictions to the total actual positives for each class, assessing the model's ability to detect card defaults.

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

**F1-Score:** Represents the harmonic mean of precision and recall, providing a balanced measure for imbalanced datasets.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{TPrecision + Recal} \quad (5)$$



The G-Mean balances the accuracy of a model on both the majority and minority classes. It is calculated as the geometric mean of sensitivity (recall) and specificity

$$G - Mean = \sqrt{Sensitivity \times Specificity} \quad (6)$$

**Specificity:** accuracy of negative examples

$$Specificity = \frac{TN}{TN + FP} \quad (7)$$

Confusion matrix: the visualization of true positives, true negatives, false positives, and false negatives.

**Area Under the Curve (AUC):** performance metric, particularly, Receiver Operating Characteristic (ROC) was used to evaluate binary classification models such as predicting whether a person will default on a loan (yes/no). These metrics ensured that the evaluation captured both class-level performance and overall predictive ability, which is particularly important given the class imbalance present in the dataset.

## 2.6. Pseudocode: Credit Card Default Prediction Using MLP with Z-Score Outlier Handling

### Input

Dataset D = (X,y)

- (a) X: Feature matrix
- (b) y: Binary target variable (0 = non-default, 1 = default)
- (c) Z-score threshold  $zth=3.0$
- (d) Train-test split ratio  $r=0.7$
- (e) Number of epochs  $E=60$
- (f) Batch size  $B=32$

### Output

- (a) Trained MLP model
- (b) Performance metrics: Accuracy, Precision, Recall, F1-score, G-mean, Specificity, AUC
- (c) Confusion matrix

### Procedure

- 1: Load dataset (X, y)
- 2: for each class  $c \in y$  do
- 3: Compute Z-scores for  $X_c$

- 4: Retain samples where  $|Z| \leq zth$
- 5: end for
- 6: Obtain cleaned dataset ( $X\_clean$ ,  $y\_clean$ )
- 7: Standardize  $X\_clean \rightarrow X\_scaled$
- 8: Split  $X\_scaled$ ,  $y\_clean$  into training (70%) and test (30%)
- 9: Define MLP with ReLU hidden layers (32, 16) and Sigmoid output
- 10: Compile model using RMSprop (learning rate: 0.0001, rho: 0.8, momentum: 0.1) and binary cross-entropy
- 11: Train model for E epochs with batch size B (20% validation)
- 12: Evaluate on test set using Accuracy, Precision, Recall, (d) F1-score, G-mean, Specificity, AUC, and confusion matrix

## 2.7 Ethical Considerations

This study adhered to ethical guidelines to ensure the integrity and fairness of the research process. The dataset was utilized solely for academic purposes and proper attribution was given to the original dataset providers where applicable. Also, the dataset was not altered in a way that could misrepresent the nature of the default, ensuring that the study remained truthful and reliable

## 3.0 Results

### 3.1. Descriptive Analysis of the Dataset

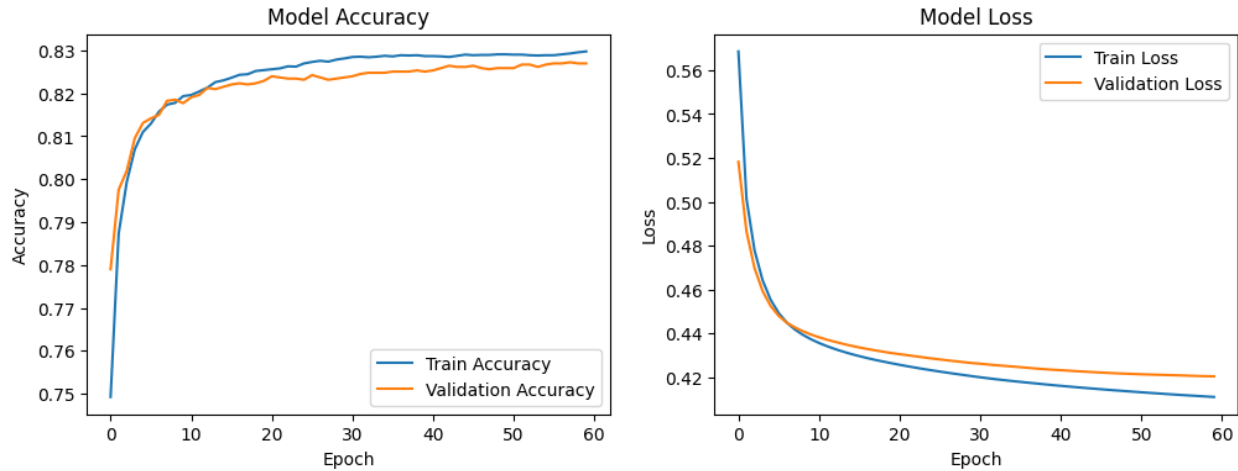
The dataset used in this study, the UCI “Default of Credit Card Clients,” consisted of 30,000 instances with 24 predictor variables. After applying Z-score outlier removal, a reduced but cleaner dataset was obtained (26, 215), ensuring that extreme and potentially misleading values were excluded from further modeling. The processed dataset was then standardized and partitioned into training and test sets using a 70:30 split.

### 3.2 Visualization of Training Process

The training of the MLP model was monitored through plots of accuracy and loss across 60

epochs as shown in Figure 3. The training and validation accuracy curves demonstrated a consistent upward trend, stabilizing at approximately 83% accuracy with minimal divergence, suggesting that the model achieved

good generalization without overfitting. Similarly, the training and validation loss curves showed a smooth decline and convergence, further confirming stable learning behavior.



**Fig 3:** Training and Validation Accuracy and Loss of the MLP Model

### 3.3 Results of the MLP Model

The performance of the MLP model on the test dataset is summarized in Table 1. The model achieved an overall accuracy of 83%, with a precision of 0.78, recall of 0.67, and F1-score of 0.70. The relatively high specificity of 0.96 indicates that the model performed very well in identifying non-defaulters. The recall, however,

at 0.67, suggests that while the model was effective at capturing defaulters, a proportion of true default cases were still misclassified. The G-mean score of 0.61 reflects this imbalance in sensitivity and specificity. The Area Under the ROC Curve (AUC) was 0.79, indicating good overall discriminatory ability between default and non-default cases.

**Table 1:** Performance Metrics of the MLP Model

Metric	Value
Accuracy	0.83
Precision	0.78
Recall	0.67
F1-score	0.70
G-mean	0.61
Specificity	0.96
AUC	0.79

### 3.4 Confusion Matrix Analysis

The confusion matrix is shown in Figure 4, including both counts and percentages presented in Table 2 illustrates the classification outcomes

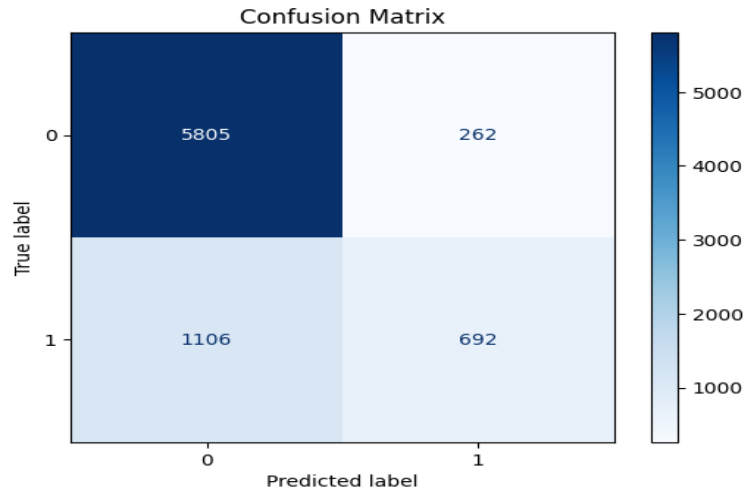
of the MLP model. Out of 6,759 non-defaulters in the test set, 5,805 were correctly classified, while 262 were misclassified as defaulters (false positives). For the 1,798 defaulters, 692 were

correctly classified, whereas 1,106 were misclassified as non-defaulters (false negatives).

The confusion matrix confirms the high specificity of the model, as the majority of non-defaulters were correctly identified. However, the relatively large number of false negatives highlights the challenge of capturing all default cases, which is a common issue in imbalanced datasets.

### 3.5 Receiver Operating Characteristic (ROC) Curve

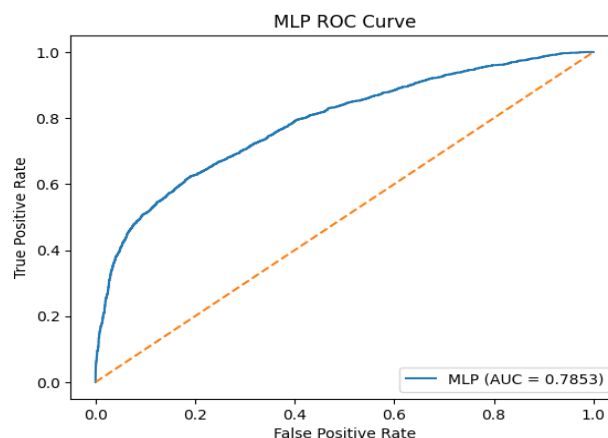
The ROC curve illustrates the trade-off between true positive rate and false positive rate presented in Figure 5. The ROC curve of the MLP model demonstrates strong discrimination between default and non-default cases, with an AUC of 0.79, indicating reliable classification performance across decision thresholds.



**Fig. 4:** Confusion Matrix of the MLP Model

Table 2: Confusion Matrix Counts and Percentages

Actual \ Predicted	Non-Default	Default
Non-Default (6,067)	5,805 (95.68%)	262 (4.32%)
Default (1,798)	1,106 (61.51%)	692 (38.49%)



**Fig. 5:** ROC curve of the MLP Model

### 3.6 Comparative Analysis

The results of the proposed MLP model with z-score outlier handling were compared against the DNN model from prior study (Bhandary &

Ghosh, 2025). Table 3 summarizes the comparative performance across multiple evaluation metrics.



**Table 3:** Comparative Performance of DNN vs Proposed MLP Model

Metric	DNN (Bhandary & Ghosh, 2025)	Proposed MLP Model
Accuracy	0.8180	0.8300
Precision	0.6390	0.7800
Recall	0.3869	0.6700
F1-score	0.4820	0.7000
G-mean	0.6027	0.6100
Specificity	0.9388	0.9600
AUC	0.7700	0.7900

The comparative analysis shows that the MLP outperformed the DNN across all metrics. The most substantial gains were observed in recall (0.67 vs. 0.39) and F1-score (0.70 vs. 0.48), suggesting that the proposed approach was considerably more effective in identifying default cases while maintaining strong precision. Improvements were also evident in accuracy (0.83 vs. 0.82), precision (0.78 vs. 0.64), specificity (0.96 vs. 0.94), and AUC (0.79 vs. 0.77). Although the G-mean improvement was marginal (0.61 vs. 0.60), the higher balance between sensitivity and specificity demonstrates robustness in handling the class imbalance characteristic of the dataset.

#### 4.0 Discussion

The findings of this study provide insights into the effectiveness of the proposed MLP model for credit card default prediction, particularly compared with a benchmark DNN. A key contribution lies in the preprocessing strategy, where Z-score-based outlier handling was applied per class. This reduced the influence of extreme observations that could disproportionately affect gradient-based learning, stabilize optimization and contribute to smoother convergence. The consistency between training and validation

curves suggests the model benefited from reduced variance rather than increased bias, supporting evidence that careful preprocessing enhances learning stability in financial datasets (Talaat et al., 2023; Dash et al., 2023; Marques et al., 2023).

Outlier removal in credit risk modeling presents an inherent trade-off. While extreme values may represent noise or data entry errors, they may also correspond to genuinely high-risk customers (Gupta, 2024; Shetabi, 2024). The Z-score method assumes approximate normality and may remove informative tail observations. The per-class implementation mitigated this risk by preserving class-specific distributions, but rare default patterns may still be excluded. This highlights the need for cautious interpretation of performance gains and suggests future work should explore alternative robust techniques, such as model-based outlier detection.

From a modeling perspective illustrated in Figure 2, the MLP model demonstrated stable learning dynamics, with accuracy and loss converging after approximately 20 epochs. The absence of significant divergence between training and validation performance indicates effective generalization, despite the relatively

shallow architecture. This challenges the assumption that deeper networks are inherently superior for credit default prediction, suggesting that architectural simplicity combined with appropriate preprocessing can sufficiently capture dominant decision patterns in structured financial data.

In terms of predictive performance, the MLP achieved 83% accuracy, with high specificity (0.96) and moderate recall (0.67). The high specificity shows the model effectively identifies non-defaulters, which minimizes unnecessary credit rejections. The moderate recall highlights the persistent difficulty in detecting default cases within an imbalanced dataset, consistent with prior literature (Yu, 2020; Jiang et al., 2022). The confusion matrix indicates errors are concentrated in false negatives rather than false positives, showing the model is conservative in flagging default risk.

Compared with the benchmark DNN, the MLP outperformed across evaluation metrics, including recall, F1-score, G-mean, and AUC. Improvements in recall and F1-score are particularly meaningful for minority-class performance and practical credit risk assessment. These gains were achieved without increasing model complexity, reinforcing that preprocessing and data representation can be as influential as network depth.

Limitations include the use of a single dataset from the Taiwanese credit market, which may constrain generalizability. Although Z-score-based outlier handling improved performance, alternative preprocessing or feature transformation strategies may further enhance predictive capability or preserve rare but informative observations. Future research should validate the approach on multiple datasets, explore cost-sensitive or threshold-adjusted learning to address false negatives, and investigate ensemble or hybrid models to

improve sensitivity without compromising specificity.

Overall, this study demonstrates that a carefully designed MLP, supported by robust preprocessing, can serve as a competitive and computationally efficient alternative to deeper neural networks for credit card default prediction. The findings highlight that predictive improvements are strongly influenced by data preparation and interpretation, not solely by model complexity.

## 5.0 Conclusion

This study addressed a critical gap in credit risk prediction: whether robust preprocessing, specifically Z-score-based outlier removal, enables simpler neural architectures such as MLPs to achieve performance comparable to or exceeding deeper models. The proposed MLP, combined with per-class outlier treatment, demonstrated superior performance across multiple evaluation metrics including accuracy, precision, recall, F1-score, and AUC, compared with a benchmark DNN. These results highlight that careful data preprocessing can be as influential as model complexity in achieving reliable predictions in imbalanced financial datasets.

Importantly, the study emphasizes the interpretability and practical applicability of the MLP model. Its relatively simple architecture, lower computational requirements, and consistent generalization performance make it well-suited for deployment in real-world credit risk systems, where transparency and efficiency are critical. By showing that a simpler, well-preprocessed model can rival deeper architectures, the study provides both methodological insights for academic research and practical guidance for financial institutions.

Future research should explore:

- Cross-validation and hyperparameter optimization to further improve model performance

- Alternative outlier detection and handling strategies to preserve rare but informative observations
- Enhanced model interpretability techniques to facilitate adoption in operational credit risk decision-making
- Statistical significance testing, such as McNemar's or paired t-tests, to formally assess model improvements over benchmarks. This is noted as future work because the original benchmark DNN data lacked distribution-level access required for such tests.

Overall, the study reinforces the value of methodological rigor, preprocessing strategies, and interpretable model design in credit risk prediction, paving the way for both theoretical advancements and practical implementation in financial applications.

### 5.1 Contribution of the Study

The study contributes to credit risk prediction in the financial sector, showing that preprocessing techniques like Z-score-based outlier removal can enhance the performance of simpler machine learning models. This approach is particularly relevant for banks and lending institutions seeking efficient, interpretable, and reliable methods for predicting credit card defaults, managing risk, and improving decision-making processes in environments with imbalanced financial datasets.

### Declarations

#### Ethics approval and consent to participate

Not applicable

#### Consent for publication

Not applicable

#### Availability of data and materials

The Default of Credit Card Clients datasets analysed during the current study is available in the UCI machine learning repository dataset <https://archive.ics.uci.edu/ml/datasets/default+f+credit+card+clients>

### Competing interests

The authors declare that they have no competing interests, regarding the research, authorship or publication of this manuscript.

### Funding

No funding was received for this study.

### Authors' contributions

R, conducted the data collection, preprocessing, model development, analysis, and interpretation of the results. A.A. served as the major supervisor, providing overall guidance, conceptual input, and critical revisions of the study. D.G. acted as the second supervisor, offering additional supervision, reviewing the methodology, critical revisions and making editorial corrections. All authors read and approved the final manuscript.

### Acknowledgements

The authors wish to thank the Department of Informatics and Secured Computing, Kaduna State University, for providing computing resources needed for this research.

### References

- Alarfaj, F. K., Malik, I., Khan, H. U., Almusallam, N., Ramzan, M., & Ahmed, M. (2022). Credit card fraud detection using State-of-the-Art machine learning and deep learning algorithms. *IEEE Access*, 10, 39700–39715. <https://doi.org/10.1109/access.2022.3166891>
- Bello, O. A. (2020). Machine Learning Algorithms for Credit Risk Assessment: An Economic and Financial analysis. *International Journal of Management Technology*, 10(1), 109–133. <https://doi.org/10.37745/ijmt.2013>
- Bhandary, R., & Ghosh, B. K. (2025). Credit Card Default Prediction: An empirical analysis on predictive performance using statistical and machine learning methods. *Journal of Risk and Financial*

- Management, 18(1), 23. <https://doi.org/10.3390/jrfm18010023>
- Blazek, R., & Duricova, L. (2025). Beyond expectations: anomalies in financial statements and their application in modelling. *Stats*, 8(3), 63. <https://doi.org/10.3390/stats8030063>
- Chang, V., Sivakulasingham, S., Wang, H., Wong, S. T., Ganatra, M. A., & Luo, J. (2024). Credit risk prediction using machine learning and Deep Learning: A study on Credit card customers. *Risks*, 12(11), 174. <https://doi.org/10.3390/risks12110174>
- Chouksey, A., Shovon, M. S. S., Tannier, N. R., Bhowmik, P. K., Hossain, M., Rahman, M. S., Rahman, M. K., & Hossain, M. S. (2023). Machine Learning-Based Risk Prediction Model for Loan Applications: Enhancing Decision-Making and Default Prevention. *Journal of Business and Management Studies*, 5(6), 160–176. <https://doi.org/10.32996/jbms.20-23.5.6.13>
- Dash, C. S. K., Behera, A. K., Dehuri, S., & Ghosh, A. (2023). An outliers detection and elimination framework in classification task of data mining. *Decision Analytics Journal*, 6, 100164. <https://doi.org/10.1016/j.dajour.2023.100164>
- Dube, L., & Verster, T. (2024). Assessing the performance of machine learning models for default prediction under missing data and class imbalance: a simulation study. *Orion/ORiON*, 40(1). <https://doi.org/10.5784/40-1-767>
- Gupta, S. (2024). The role of artificial intelligence in predicting credit risk. *International Journal of Artificial Intelligence Data Science and Machine Learning*, 5. <https://doi.org/10.63282/-3050-9262.ijaidsm-l-v5i4p114>
- Han, Y., & Joe, I. (2025). Enhanced predictive modeling for anomaly detection in financial transactions using machine learning. *IEEE Access*, 13, 154438–154449. <https://doi.org/10.1109/access.-2025.3602236>
- James, O. (2023). *The impact of financial technology on loan risk and payment systems: Evidence from Africa - UEA Digital Repository*. University of East Anglia. Norwich Business School. <https://ueaeprints.uea.ac.uk/id/eprint/95869>
- Jiang, J., Meng, X., Liu, Y., & Wang, H. (2022). An enhanced TSA-MLP model for identifying credit default problems. *SAGE Open*, 12(2). <https://doi.org/10.1177/21582440221094586>
- Lartey, C., Liu, J., Asamoah, R. K., Greet, C., Zanin, M., & Skinner, W. (2024). Effective outlier detection for ensuring data quality in flotation data modelling using machine learning (ML) algorithms. *Minerals*, 14(9), 925. <https://doi.org/10.3390/min14090925>
- Mestiri, S. (2024). Credit scoring using machine learning and deep Learning-Based models. *Data Science in Finance and Economics*, 4(2), 236–248. <https://doi.org/10.3934/dsfe.2024009>
- Nkambule, D. S., Twala, B., & Pretorius, J. H. C. (2024). Effective Machine Learning Techniques for Dealing with Poor Credit Data. *Risks*, 12(11), 172. <https://doi.org/10.3390/risks12110172>
- Shetabi, M. (2024). Evolutionary-based ensemble feature selection technique for

- dynamic application-specific credit risk optimization in FinTech lending. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-024-06369-8>
- Shu, X., & Ye, Y. (2022). Knowledge Discovery: Methods from data mining and machine learning. *Social Science Research*, 110, 102817. <https://doi.org/10.1016/j.ssresearch.2022.102817>
- Talaat, F. M., Aljadani, A., Badawy, M., & Elhosseini, M. (2023). Toward interpretable credit scoring: integrating explainable artificial intelligence with deep learning for credit card default prediction. *Neural Computing and Applications*, 36(9), 4847–4865. <https://doi.org/10.1007/s00521-023-09232-2>
- Yu, Y. (2020). The application of machine learning algorithms in credit card default prediction. & qout; 2020 *International Conference on Computing and Data Science (CDS)*, Stanford, CA, USA, 2020. <https://doi.org/10.1109/cds4970-3.2020.00050>