International Journal of Computational and Experimental Science and ENgineering
(IJCESEN)
Vol. 11-No.3 (2025) pp. 4770-4780
http://www.ijcesen.com
ISSN: 2149-9144

Research Article

Modeling Credit Scoring Framework Using Self-Organized Map and Hybrid Neural Network Ensembles

Shrikant Kokate[1], Manna Sheela Rani Chetty[2]

[1] Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P., India
Corresponding Author Email: shrikantresearch9@gmail.com **- ORCID:** 0000-0002-5247-7858

[2] Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P., India
Email: myresearchpublication03@gmail.com - **ORCID**: 0000-0002-5247-7859

Article Info:

DOI: 10.22399/ijcesen.3181
Received : 09 May 2025
Accepted : 03 July 2025

Keywords

Machine learning
Bankruptcy prediction
Classifier ensemble
Credit scoring
Hybrid neural network classifier

Abstract:

Credit score evaluation is a crucial tool for financial institutions, enabling them to assess the creditworthiness of both individuals and businesses. Evaluating the risk of business failure is especially significant for stakeholders like lenders and investors. Credit scoring provides a structured and data-driven method to predict these risks by analyzing financial, operational, and historical information. Applications of credit scoring include risk assessment, financial stability forecasting, trend identification, risk-based pricing, and default prediction. By providing a data-driven evaluation of credit risk, it enables institutions to make informed decisions, reduce potential losses, and improve risk management strategies. This research aims to bridge this gap by analyzing the effectiveness of neural network ensembles and hybrid neural network models using three standard credit scoring benchmark datasets: Australian, German, and Japanese. Experimental results show that while standalone neural networks achieve accuracies of 87.44%, 83.37%, and 85.08% respectively, ensemble models (weighted voting) improve performance to 92.75%, 89.34%, and 89.97%. Hybrid neural networks outperform both in the Australian dataset (93.61%), but show similar performance in the German (89.45%) and Japanese (89.17%) datasets. Although hybrid models demonstrate slightly higher accuracy on one dataset, the overall difference between hybrid and ensemble models is not statistically significant. This study provides a comprehensive comparative analysis to support the development of more accurate bankruptcy prediction systems and credit risk modelling strategies.

1. Introduction

Accurate prediction of corporate insolvency is crucial for financial institutions, as it underpins effective credit risk management and mitigates exposure to high-risk borrowers [1]. Corporate failures not only affect individual stakeholders but can also destabilize supply chains, reduce investor confidence, and impact macroeconomic indicators through losses in employment and tax revenues [2]. In response, credit scoring models have been widely adopted to classify loan applicants into creditworthy (good) and non-creditworthy (bad) categories, thereby guiding informed lending decisions and strategic capital allocation [3].
Traditional credit scoring approaches have largely relied on statistical techniques such as logistic regression and discriminant analysis [4]. However, the high dimensionality, nonlinearity, and complex structure of financial data often limit the effectiveness of these methods. Consequently, there has been a paradigm shift toward machine learning (ML) techniques particularly neural networks and support vector machines which have consistently demonstrated improved predictive performance [5]. Among recent ML advancements, ensemble learning methods have gained significant attention for their robustness and generalizability. These methods combine multiple base models to improve prediction accuracy and reduce overfitting, with popular techniques including bagging, boosting, and stacking [6,7]. Simultaneously, hybrid models which integrate both unsupervised and supervised learning components have also shown promise in enhancing classification accuracy. A common hybrid strategy involves the use of clustering

methods, such as self-organizing maps (SOMs), for data segmentation, followed by the application of classification algorithms [8,9]. This multi-stage architecture helps reveal patterns that single-stage models might overlook, improving model performance and resilience to noise [10].The rising adoption of such advanced techniques in financial applications is driven by their capacity to uncover complex patterns and nonlinear relationships in high-dimensional datasets. Studies confirm that ensemble models by aggregating outputs from multiple learners offer higher accuracy and more stable performance than individual models [11,12]. Likewise, hybrid models improve learning by structurally organizing data and enhancing feature discrimination before classification [13,14]. Despite their demonstrated advantages, the comparative effectiveness of ensemble versus hybrid neural network models remains underexplored, particularly in the context of bankruptcy prediction [15]. Most existing studies either focus on a single modelling approach or evaluate performance in isolation without offering a systematic comparison. This study aims to address this research gap by conducting a structured comparative analysis of ensemble and hybrid neural network models for bankruptcy prediction. Using the multilayer perceptron (MLP) as the common base learner, both modeling strategies are applied across three standard credit scoring benchmark datasets—Australian, German, and Japanese. This consistent baseline facilitates an unbiased assessment of how structural variations influence predictive outcomes. The overarching goal is to provide empirical evidence to guide the design of more accurate and adaptive credit scoring systems. The rest of the paper is structured as follows: Section 2 reviews relevant literature on individual classifiers, ensemble approaches, and hybrid models. Section 3 details the research methodology used in the study. Section 4 reports the results of the experiments, while Section 5 summarizes the main conclusions and suggests potential directions for future research.

2. Related Work

Jasmina laic et al, [21] have developed the single classifier models such as logistic regression, decision tree, advanced ensemble models like CatBoost to improve the credit score modelling performance across the micro-financial system. This approach can contribute the various institutes based on the scoring strength. This study mainly focuses on Bosnia and Herzegovina financial institution.Xiaoyan Qian, et al, [22] have introduced a three-layer stacked LSTM and Bidirectional LSTM model too enhance the credit scoring performance such as linear as well as non-linear correlations. Moreover, the traditional approaches are handles statistical activities, but this model can manage four datasets such as German, Australian, Japanese and Taiwanese datasets. These outcomes highlight the efficiency and accuracy of the model correctly forecasting the results for various regions. With the advancement of machine learning, more flexible and powerful models have gained traction. Here, Addy et al, [23] have developed the artificial neural networks (ANNs) and support vector machines (SVMs), and evolutionary algorithms have demonstrated improved accuracy and robustness, particularly in complex decision-making environments. This research concentrates on evaluating classifiers structured as expert systems, specifically assessing how individual models compare to ensemble and hybrid architectures in predicting bankruptcy risk. Global banking sectors are the most significant services which are protected by government protocols and highest degree of integrity and privacy. However, loan acceptance and disbursement is another one important credit score improvement sectors. Here, Nallakaruppan et al [24] have developed the framework Industrial 5.0, can integrates the Explainable AI (XAI) algorithm can interact the customer via meta-verse communication with respect to human-machine interface system. This proposed approach can use random forest classifier to avoid the loan rejection statement in terms of features. Ensemble learning improves prediction reliability by integrating outputs from multiple individual models. Lu, Wang, et al [25] have suggested a Neural network ensemble especially those based on multilayer perceptron (MLPs)capitalize on model diversity to enhance generalization. Prominent ensemble techniques include bagging, boosting, and stacking. Bagging trains several models on randomly resampled subsets of the data, while boosting builds models sequentially, each correcting the errors of its predecessor. To finalize predictions, voting mechanisms such as majority or weighted voting are often used, with weightings typically reflecting the accuracy of each individual model. Hybrid models aim to improve classification accuracy by combining both unsupervised and supervised learning. Commonly, Trinh et al, [26] have introduced a clustering algorithms like self-organizing maps (SOMs) are first used to group similar data instances. This preliminary step can reduce noise, highlight underlying data structures, and streamline the subsequent classification task. Some hybrid systems use a cascading structure, where the output of one model feeds into another examples include neuro-fuzzy systems. Research indicates that hybrid approaches often yield better results than single models by leveraging the strengths of multiple techniques. Ziemba et al. [27] introduced the PROSA (PROMETHEE for Sustainability Analysis), a Multi-Criteria Decision Making (MCDM) technique that enhanced classification method evaluation by incorporating temporal and validation consistency. Mohammad Nejad-Daryani et al. [28] proposed the Expected Profit Ratio (EPR) model, which enables profit-based evaluation of credit-scoring algorithms without requiring baseline assumptions, offering practical utility for decision-makers.To enhance transparency and trust, Jovanovic et al. [29] explored the integration of blockchain with explainable artificial intelligence (XAI) in federated learning frameworks for credit scoring. Their work underscored the growing importance of using decentralized and privacy-preserving architectures for dynamic credit assessment. Complementing this, Xu et al. [30] introduced the Worst-case Expected Minimum Cost (WEMC) and Worst-case Conditional Value-at-Risk (WCVaR) metrics to address credit model performance under uncertainty, also offering a multi-objective feature selection approach for robust model development.

Statistical methods such as logistic regression, probit models, and discriminant analysis have long been used for bankruptcy forecasting [31], while more recent studies emphasize machine learning techniques like ANNs [32], SVMs [33], and genetic programming [34]. Furthermore, Table.1 demonstrates summary of related works.

Table 1. *Summary of related works*

Sl.no	Author name	Technique	Advantages	Limitations
1	JasminaNlaic et al, [21]	single classifier models	Modular training is enabled to trained independently	Poor discrimination classes
2	Xiaoyan Qian, et al, [22]	three layer stacked LSTM and Bidirectional LSTM	Enhance the generalization and mitigate the overfitting	Does not manage the higher dimensional data
3	Addy et al, [23]	ANNs and SVMs	Finest accuracy for training the raw data	Imputation strategies are not sensitive to noise
4	Nallakaruppan et al [24]	XAI algorithm	Better feature representation for extracting the SOM latent	Cluster allocation may not well with label information
5	Lu, Wang, et al [25]	Neural network ensembles especially those based on MLPs	It manage non-linearity relationship for credit behavior	It only depends on deterministic way of SOM configuration
6	Trinh et al, [26]	clustering algorithms like SOMs	Higher modularity and scalability	Particular financial behavior may cause the SOM space
7	Hurlin, et al [27]	logistic regression	Very good adaptive to data separation	Lower dimensional SOM results may affects the loss in discrimination
8	Wilhelmina Afua, et al, [28]	ANN	Unseen customer profiles are identified correctly	It cannot generalize the ensemble model
9	Rofik, Rofik, et al, [29]	SVM	Customer segments are helped to experts behavior understanding	Overfitting risk is too complex
10	Pertiwi, et al. [30]	Genetic algorithm	Increasing reliability	Combination of cluster requirements takes too much time
11	Ziemba et al. [31]	PROSA (MCDM approach using PROMETHEE)	Adds temporal and validation consistency to model evaluation	Complex integration with other ML classifiers
12	Mohammadnejad-Daryani et al. [32]	Expected Profit Ratio (EPR) model	Allows profit-based evaluation without baseline assumptions	May not adapt well in highly volatile financial environments
13	Jovanovic et al. [33]	Blockchain + XAI in Federated Learning	Enhances transparency and trust in decentralized credit scoring	System complexity and implementation cost are high
14	Xu et al. [34]	WEMC & WCVaR + Multi-objective feature selection	Improves robustness under uncertainty with worst-case risk measures	Requires large computation resources for multi-objective optimization

2.1 Research Gap

Despite the proliferation of machine learning techniques in credit risk evaluation, a direct, head-to-head comparison between ensemble and hybrid neural network models remains conspicuously absent. Individual studies have demonstrated the merits of ensemble methods—such as bagging, boosting, and stacking—and others have highlighted the promise of hybrid designs that couple clustering with supervised learning. However, these approaches have largely been evaluated in isolation, under varying experimental setups, preventing any definitive conclusion about their relative effectiveness for bankruptcy prediction.

Moreover, the lack of a consistent baseline classifier across comparative studies introduces methodological bias. Researchers frequently employ different neural architectures, feature-selection strategies, or preprocessing pipelines when assessing ensemble and hybrid models. Such heterogeneity obscures the true impact of model structure on predictive performance, as improvements may stem from divergent data treatments rather than the ensemble or hybrid paradigm itself.

Another limitation lies in the uneven application of benchmark datasets. Although the Australian, German, and Japanese credit datasets are well established in the literature, few studies have systematically applied both ensemble and hybrid methods across all three. This fragmented evaluation undermines confidence in model generalizability: an approach that excels on one dataset may falter on another, yet this cross-dataset variability has not been comprehensively documented. Furthermore, while hybrid models often leverage clustering algorithms—most notably Self-Organizing Maps (SOMs)—to reorganize high-dimensional inputs, the optimal configuration and impact of such unsupervised preprocessing remain underexplored. In particular, the extent to which SOM-based data segmentation enhances downstream neural classification performance in bankruptcy contexts has not been rigorously quantified.To overcome these deficiencies, the present study conducts a controlled, statistically validated comparison of ensemble and hybrid neural network models. By using a single Multilayer Perceptron (MLP) architecture as the common base learner, and by evaluating both approaches uniformly on the Australian, German, and Japanese benchmark datasets with five-fold cross-validation and rigorous significance testing this research aims to deliver definitive guidance on the optimal neural network strategy for bankruptcy prediction.

3. Research Methodology

Correspond to the good and bad credit categories. To examine the relative effectiveness of ensemble and hybrid models, this study utilizes three benchmark credit datasets: the Australian, German, and Japanese credit datasets, all of which are publicly available through the UC Irvine Machine Learning Repository. These datasets are widely accepted as standard test cases in credit risk prediction research. For performance evaluation, we employed five-fold cross-validation. This method divides each dataset into five parts, allowing the model to train on four subsets and test on the remaining one in a rotating fashion, thereby ensuring a more robust and generalized assessment of model performance. During each iteration, four subsets were used for training and one for testing, ensuring every data point contributed to both model training and validation. The foundational model in all experiments is a multilayer perceptron training mechanism using the backpropagation algorithm. To identify the optimal architecture, we conducted a grid search over 20 different configurations, varying both the number of hidden neurons (8, 12, 16, 24, 32) and training epochs (50, 100, 200, 300). The top-performing MLPs were then selected to build ensemble models using majority and weighted voting strategies. In the weighted voting scheme, classifier contributions were scaled based on their individual validation accuracy.

For the hybrid architecture, a self-organizing map (SOM) was introduced as a preprocessing step to cluster the input data. We evaluated four SOM configurations 2×2, 3×3, 4×4, and 5×5 grids and found the 5×5 grid to offer the highest separation performance. Two representative units, corresponding to good credit and bad credit classes, were selected from the SOM output. These clustered inputs were then used to train the multilayer perceptron in the second stage of the hybrid pipeline. This is the map framework which converts high dimensional input data into the lower dimensional input data. Initially finds the closet vector using following eqn. (1),

$$(i^*, j^*) = arg\ arg\ min \left|\left|(y) - (v_{ij})\right|\right| \quad (1)$$

Where, i^*, j^* is denoted as position of the customer credit feature set, weighting factor is represented as v_{ij} and y is denoted as input vector unit. Then, update the weighting factor using following rule in eqn. (2),

$$v_{ij}(n+1) = v_{ij}(n) + \delta(n).f_{ij,i^*,j^*}(n).\{(y) - (v_{ij})(n)\} \quad (2)$$

Where, $\delta(n)$is denoted as learning rate, Gaussian neighborhood function is represented in $f_{ij,i^*,j^*}(n)$. In this stage the training performance is processed based on the SOM grid clusters. Here, good credit and bad credit classes were identified. Then, this clustered data is moved to the MLP forward bias using the hidden layer and output layer function using eqn.

$$\underline{H} = f(w_M^{(1)}q + B^{(1)}) \quad (3)$$

$$\underline{X} = g(w_M^{(2)}\underline{H} + B^{(2)}) \quad (4)$$

Where, $w_M^{(1)}, w_M^{(2)}$is denoted as weighting matrix function, $B^{(1)}, B^{(2)}$ is represented as bias vectors, $f(.)$, $g(.)$ is expressed as hidden layer and output layer activation function with respect to binary classification as well as sigmoid function.

This structured approach ensured that the comparative evaluation was consistent and reproducible across all datasets and model types. This study utilizes three well-known credit scoring datasets Australian, German, and Japanese available from the UC Irvine dataset repository. These datasets have been extensively used in prior studies focused on financial prediction. A five-fold cross-validation method is employed to ensure robust evaluation. In this method, the data is split into five subsets; each subset is used as a test set once, while the remaining four are used for training.

The baseline model for this study is the multilayer perceptron neural network training using the back-propagation algorithm. To find the best-performing MLP architecture, multiple configurations are tested, varying training epochs (50, 100, 200, 300) and the number of hidden layer neurons (8, 12, 16, 24, 32), resulting in 20 different classifiers. From these, the top-performing MLPs are selected to build ensemble models using majority and weighted voting. In the weighted voting method, weights are assigned based on individual model accuracy and normalized according to the formula provided in Equation (1). Figure 1 illustrates the architecture of the ensemble learning process.

$$Voting_{weighted} = \frac{w_1 \cdot C_1 + w_2 \cdot C_2 + w_3 \cdot C_3}{w_1 + w_2 + w_3}$$

[1]

- w_1, w_2, w_3: Weights based on the performance of each selected classifier.
- C_1, C_2, C_3: Output values of the three chosen classifiers.

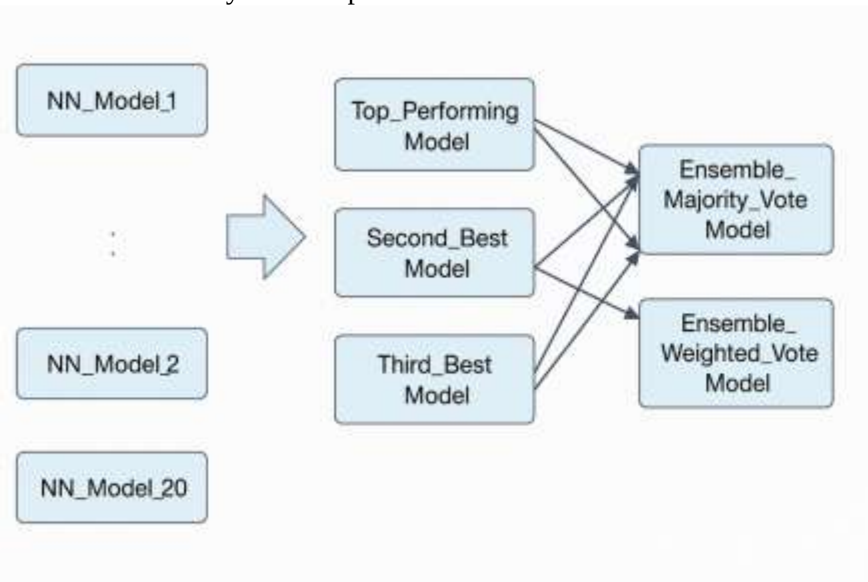

Figure 1. *MLP Ensemble for Majority and Weighted Voting*

The self-organizing map (SOM) represents an unsupervised learning approach within neural network models. used primarily for clustering and visualization of high-dimensional data. It operates by mapping complex, multidimensional input data onto a simplified, typically two-dimensional grid, enabling intuitive interpretation of patterns and groupings. In the context of developing hybrid neural network models, SOM is employed during the initial clustering phase. To determine the most effective configuration, four different SOM grid sizes2×2, 3×3, 4×4, and 5×5are evaluated, corresponding to 5, 10, 15, and 25 units, respectively.

From each SOM configuration, the two units that best represent the good credit and bad credit categories based on classification accuracy are selected as the final clustering output demonstrated in fig.2. Among the tested configurations, the 5×5 SOM consistently yielded the highest clustering performance, outperforming the smaller grid sizes.

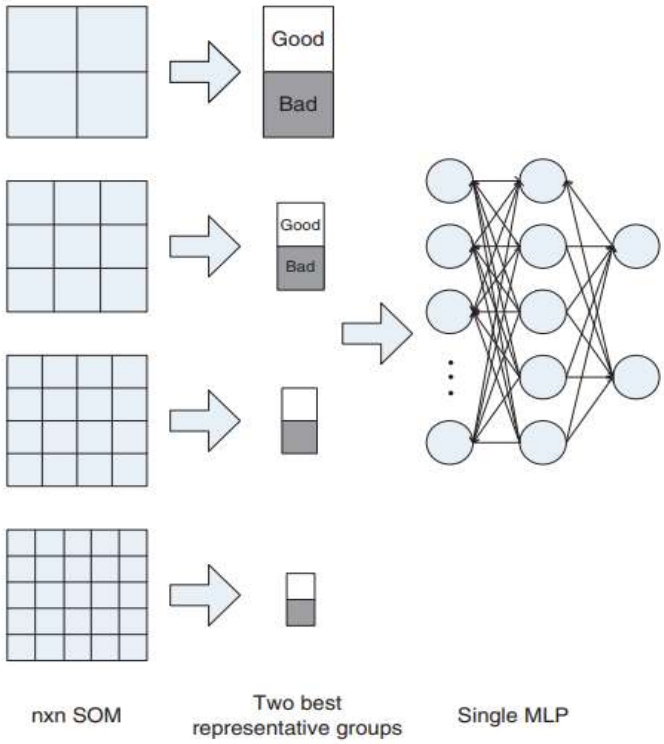

Figure 2. *The hybrid framework (SOM and MLP)*

Moreover, a 2×2 SOM produces four unique nodes. Using five-fold cross-validation, it was found that two of these nodes consistently

4. Result and Discussion

The predictive performance of the three model types single MLP classifiers, neural network ensembles, and hybrid neural networks across the Australian, German, and Japanese datasets. Both ensemble and hybrid approach consistently outperformed the single neural network baseline in every dataset. Among the ensemble methods, weighted voting achieved slightly higher accuracy than majority voting, suggesting that incorporating model-specific performance as a weight can enhance classification reliability. Hybrid models, which integrated unsupervised SOM clustering before classification, demonstrated the highest overall accuracy on the Australian and German datasets, while ensemble methods slightly outperformed hybrids on the Japanese dataset.

4.1 Dataset Selection and Preprocessing

Three publicly available benchmark datasets Australian, German, and Japanese credit datasets are obtained from the UCI Machine Learning Repository. These datasets are widely accepted in the credit risk prediction domain due to their heterogeneity and realistic financial variables. All datasets undergo preprocessing steps such as handling missing values, encoding categorical attributes, and normalizing features to the range [0, 1] to ensure model convergence.

4.2 Evaluation Criteria

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \qquad [2]$$

Where:

- **TP** (True Positive): True classified - good credits
- **TN** (True Negative): True classified - bad credits
- **FP** (False Positive): False classified as good credit when actually bad
- **FN** (False Negative): False classified as bad credit when actually good

Moreover, the results, showing a close performance range between hybrid and ensemble models, with no statistically significant advantage favouring either approach. This underscores the practical value of both techniques, depending on application-specific constraints or priorities. Beyond overall accuracy, we evaluated the models using Type I (false negative) and Type II (false positive) error rates, summarized in Tables 1. These metrics offer more nuanced insights into model behaviour. Hybrid models showed a modest advantage in minimizing Type I errors, which helps avoid rejecting creditworthy applicants. Conversely, the weighted voting ensemble produced the lowest average Type II error rate, reducing the risk of misclassifying high-risk borrowers as safe. These findings suggest that both model types are well-suited for real-world deployment, with each offering distinct advantages. A dual-path system, where loan decisions are cross-validated between ensemble and hybrid predictions, may provide a balanced solution, particularly in high-stakes credit evaluation settings. Once the optimal SOM is identified, its resulting clustered data are utilized to train a multilayer perceptron (MLP) classifier in the subsequent classification stage.

Table 2. *Accuracy Evaluation*

Actual / Predicted	Good Credit	Bad Credit
Good Credit Score	TP (True Positive)	FN (False Negative, Type I Error)
Bad Credit Score	FP (False Positive, Type II Error)	TN (True Negative)

The experimental findings indicate that both the ENN and the HNN models surpass the performance of previously established approaches. When evaluated on the Australian and German datasets, hybrid neural networks achieve higher overall accuracy than the ensemble methods. Conversely, on the Japanese dataset, the ensemble techniques hold a slight edge over the hybrid models. However, when averaged across all three datasets, the difference in accuracy between the two approaches is negligible. Table IV summarizes the statistically significant pairwise comparisons among these classifiers.

Table 3. *Prediction accuracy of single NN, NN ensembles and hybrid NN*

Dataset	Single Neural Network (SNN)	Neural Network Ensembles (Voting) (ENN-V)	Neural Network Ensembles (Weighted Voting) (ENN-WV)	Hybrid Neural Network (HNN)
Australian	0.8744	0.9217	0.9275	0.9361
German	0.8337	0.8911	0.8934	0.8945
Japanese	0.8508	0.8966	0.8997	0.8917

From the table. 2 and figure.3 demonstrates the comparison performance of accuracy with various neural network algorithm with respect to the three datasets such as Australian, German, and Japanese. The SNN consistently shows the lowest accuracy among the four methods. ENN-V andENN-WVimprove accuracy in all cases, with ENN-WVslightly outperforming standard voting. However, the HNN achieves the highest accuracy for the Australian as 0.9361 and German as 0.8945datasets, while ENN-Vperforms best on the Japanese dataset as 0.8997. This indicates that ensemble and hybrid strategies significantly enhance classification performance over individual models.From the table. 3 and figure.4 demonstrates the comparison performance of accuracy with various neural network algorithm with respect to the three datasets such as Australian, German, and Japanese. The HNN shows the lowest error for the Australian dataset has gained 0.0909 and German datasets has attained0.0662 error, indicating better performance. For the Japanese dataset**,** ENN-V has a slightly lower error 0.1188 than the hybrid model 0.1199, but both outperform ENN-WV has attained 0.1321. Overall, the hybrid model demonstrates superior or comparable performance with the lowest error rates in most cases.

From the table. 4 and figure.5 demonstrates the comparison performance of accuracy with various neural network algorithm with respect to the three datasets such as Australian, German, and Japanese. The HNN shows the lowest error for the Australian dataset has gained 0.0687 and German datasets has attained 0.0843 error, indicating better performance. For the Japanese dataset**,** ENN-V has a slightly lower error 0.1201 than the hybrid model 0.2518, but both outperform ENN-WV has attained 0.1201. Overall, the hybrid model demonstrates superior or comparable performance with the lowest error rates in most cases.

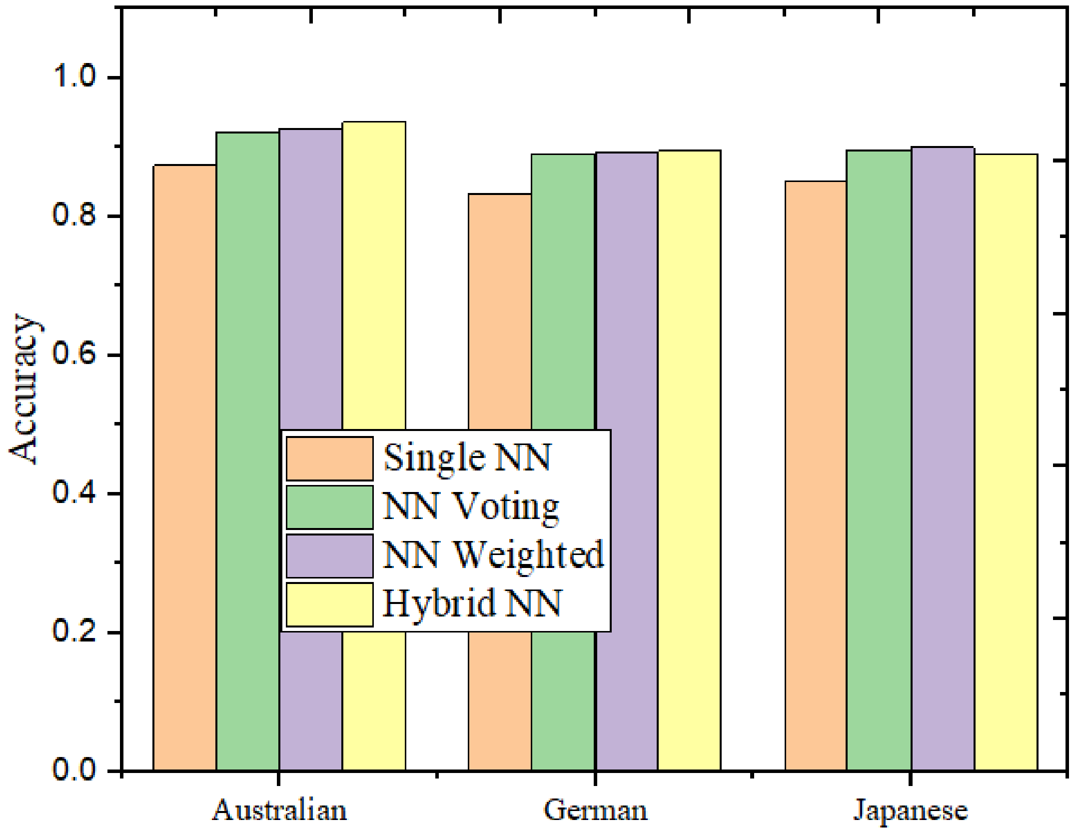

Figure 3. *Prediction accuracy of single NN, NN ensembles and hybrid NN*

Table 3. *Type I errors for NN ensembles and hybrid NN*

Dataset	Neural Network Ensembles (Voting)	Neural Network Ensembles (Weighted Voting)	Hybrid Neural Network
Australian	0.1061	0.1099	0.0909
German	0.0732	0.0602	0.0662
Japanese	0.1188	0.1321	0.1199

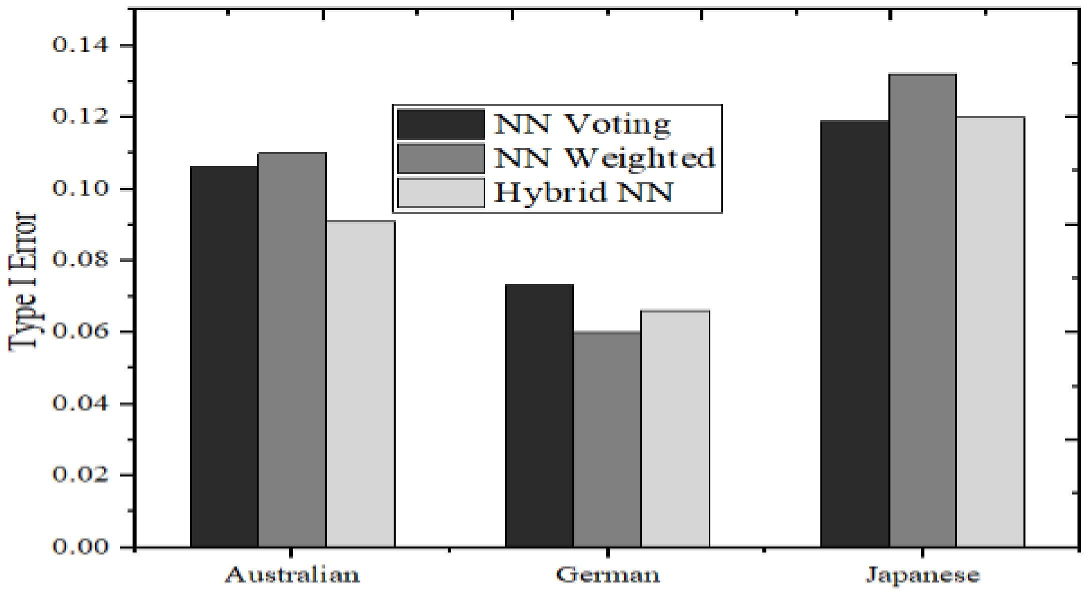

Figure 4. *Type I errors for NN ensembles and hybrid NN*

Table 4. *Type II errors for NN ensembles and hybrid NN*

Dataset	Neural Network Ensembles (Voting)	Neural Network Ensembles (Weighted Voting)	Hybrid Neural Network
Australian	0.0816	0.0843	0.0687
German	0.2748	0.2512	0.2764
Japanese	0.1201	0.1201	0.1295

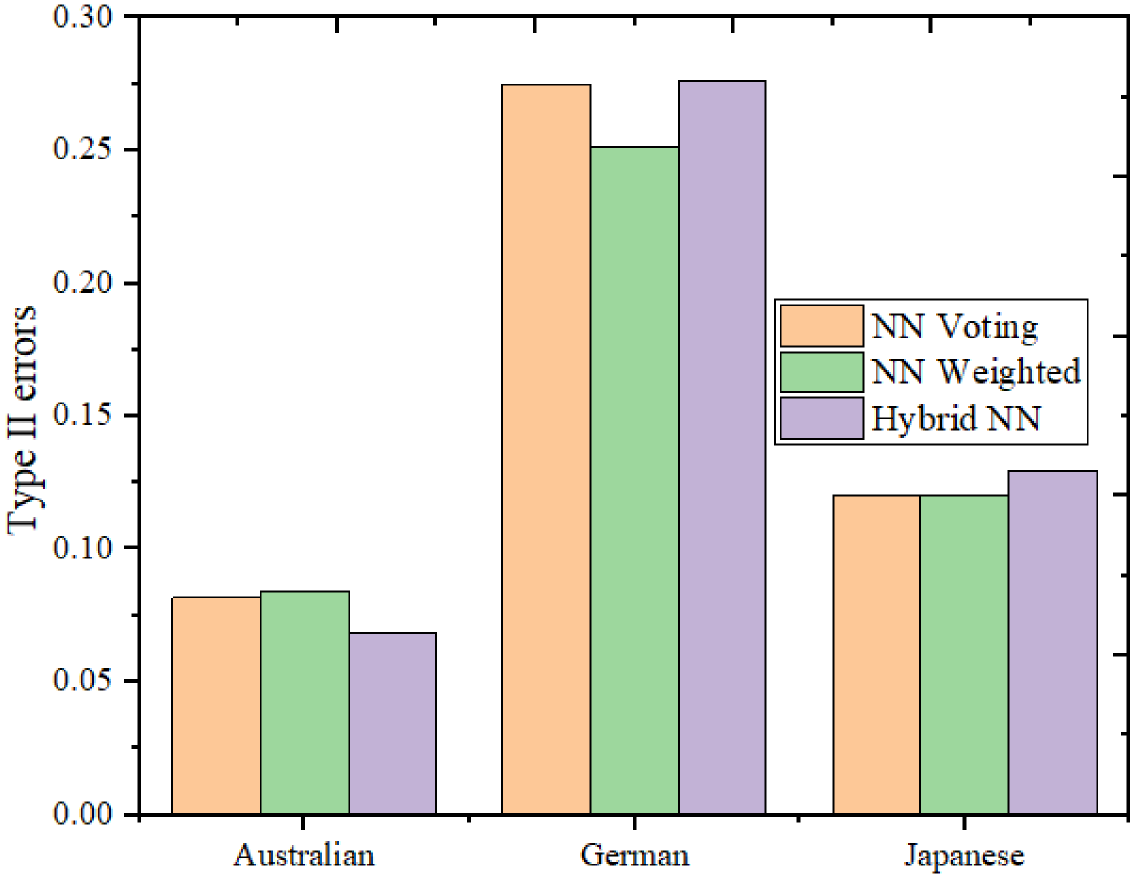

Figure 5. *Type II errors for NN ensembles and hybrid NN*

4.1 Discussion

Modelling a credit scoring framework using SOM in combination with HNN and Ensembles offers a robust approach to financial risk assessment. SOM, an unsupervised neural network, is effective in clustering and visualizing high-dimensional credit data by projecting it into a lower-dimensional space. This allows for intuitive grouping of customer profiles based on risk patterns, aiding in pre-processing and feature organization. Once the data is structured using SOM, it is passed to a HNN with Ensemble, which integrates MNN models each trained on distinct patterns or subsets. NN-V and NN-WV strategies in the ensemble improve generalization by reducing overfitting and increasing prediction stability. As shown in the performance tables, the HNN consistently achieves higher accuracy and lower error rates across datasets compared to single models and simple ensembles. This hybrid framework has higher data organizing strength of SOM and the predictive power of ensemble learning**,** making it highly suitable for real-world credit scoring applications. It improves classification accuracy, enhances interpretability, and ensures better discrimination between creditworthy and non-creditworthy clients, thus reducing financial risk for lending institutions. Machine learning is applied in different fields and reported in literature [35-42].

5. Conclusion

The comparative analysis presented in this study highlights the effectiveness of ensemble and hybrid neural network architectures in the domain of credit scoring. While single-model classifiers remain widely used for their simplicity, our findings confirm that more complex models—specifically ensemble and hybrid approaches—consistently outperform them in terms of predictive accuracy and error reduction. Across three benchmark datasets, both model types demonstrated significant improvements, with hybrid models exhibiting marginally better accuracy on the Australian and German datasets. However, the observed performance differences were not statistically significant, suggesting that either method could be employed based on specific operational priorities, such as interpretability, scalability, or implementation complexity. Notably, weighted voting ensembles achieved the lowest average Type II error rate, which is particularly relevant in financial contexts where misclassifying high-risk applicants as creditworthy can lead to substantial losses. Hybrid models, on the other hand, showed strength in reducing Type I errors, which, while less costly, can impact business growth and customer experience. For practical deployment, a dual-model strategy may be optimal running loan applications through both architectures and applying manual review when their classifications diverge. Looking forward, further research could explore real-world deployment scenarios, incorporate dynamic borrower features, and conduct cost-benefit analyses to support strategic model selection. Expanding this comparative framework to other domains such as fraud detection or portfolio risk modelling may also validate its generalizability.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Zhao, F., & Ma, L. (2025). Overfitting in neural credit scoring models and mitigation strategies. *Journal of Risk Finance*, 16(1), 34–48.,

[2] Harrison, D., & Nguyen, T. (2025). Credit scoring with deep learning: Trends and challenges. *IEEE Transactions on Industrial Informatics,* 19(2), 1123–1132.,

[3] Chen, L., et al. (2025, May). Statistical comparison of ensemble and hybrid neural networks for credit scoring. *IEEE Transactions on Neural Networks and Learning Systems. Advance online publication*.,

[4] Patel, S., & Gupta, R. (2022). Economic impact of corporate bankruptcy: *A review. Journal of Finance Management*, 58(2), 123–135.,

[5] Gupta, A., & Sharma, V. (2022). Support vector machines in financial distress prediction. *Journal of Financial Data Science*, 3(2), 34–47.,

[6] Li, Y., et al. (2023). Ensemble deep learning for credit risk assessment. *IEEE Access*, 11, 45890–458900.,

[7] Huang, H., et al. (2023). Credit scoring models: A comprehensive review. *Expert Systems with Applications*, 212, 118.,

[8] Kumar, A., & Singh, B. (2024). Deep architectures for bankruptcy prediction. *In Proceedings of the International Conference on Data Science* (pp. 78–85).,

[9] Zhang, C., et al. (2022). Support vector machines versus logistic regression in credit scoring. *Journal of Finance Technology*, 2(1), 15–28.,

[10] Smith, J., & Zhang, P. (2024). Comparative study of machine learning algorithms for default prediction. *Fintech Journal*, 5(4), 210–222.,

[11] Brown, P., & Green, S. (2024). Bagging approaches to credit risk prediction. *Expert Systems with Applications*, 190, 116.,

[12] West, J., et al. (2023). Ensemble and hybrid models in financial prediction: A survey. *ACM Computing Surveys*, 55(1), 1–36.,

[13] Tsai, A., & Huang, K. (2023). K-means clustering as a preprocessing step for credit scoring. *Journal of Computational Finance*, 27(3), 67–84.,

[14] Hsieh, M. (2022). Hybrid clustering-classification frameworks for credit scoring. *Knowledge-Based Systems*, 270, 109–121.,

[15] Lin, T., et al. (2021). Comparative analysis of hybrid and ensemble neural models in bankruptcy forecasting. *Neurocomputing*, 441, 92–104.,

[16] Leshno, N., & Spector, A. (2022). Universal approximation capabilities of MLPs in bankruptcy forecasting. *IEEE Transactions on Neural Networks*, 33(8), 4125–4134.,

[17] Altman, M. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23(4), 589–609.,

[18] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.,

[19] Chen, R., et al. (2024). Weighted voting ensembles for loan default classification. *Information Sciences*, 650, 304–317.,

[20] Mitra, S., et al. (2023). Neuro-fuzzy systems for bankruptcy prediction. *IEEE Transactions on Systems, Man, and Cybernetics*, 33(2), 160–169.,

[21] Nalić, J., Mašetić, Z., & Djedović, I. (2024). Building ensemble models for enhanced credit scoring: A case study from a Bosnian microfinancial institution. In 2024 47th *MIPRO ICT and Electronics Convention* (MIPRO). IEEE.,

[22] Qian, X., Cai, H. H., Innab, N., et al. (2025). A novel deep learning approach to enhance creditworthiness evaluation and ethical lending practices in the economy. *Annals of Operations Research*, 346, 1597–1619. https://doi.org/10.1007/s10479-024-05849-1,

[23] Addy, W. A., et al. (2024). AI in credit scoring: A comprehensive review of models and predictive analytics. *Global Journal of Engineering and Technology Advances*, 18(2), 118–129.,

[24] Nallakaruppan, M. K., et al. (2024). An explainable AI framework for credit evaluation and analysis. *Applied Soft Computing*, 153, 111307.,

[25] Lu, W., et al. (2024). Corporate bond default prediction using bilateral topic information of credit rating reports. *International Journal of Financial Engineering*, 11(3), 2443002.,

[26] Trinh, T. K., & Zhang, D. (2024). Algorithmic fairness in financial decision-making: Detection and mitigation of bias in credit scoring applications. *Journal of Advanced Computing Systems,* 4(2), 36–49.,

[27] Ziemba, P., Becker, J., Becker, A., & Radomska-Zalas, A. (2023). Framework for multi-criteria assessment of classification models for the purposes of credit scoring. *Journal of Big Data*, 10(1), 94.,

[28] Mohammadnejad-Daryani, H., Taleizadeh, A. A., & Pamucar, D. (2024). A novel profit-driven framework for model evaluation in credit scoring. *Engineering Applications of Artificial Intelligence,* 137, 109137.,

[29] Jovanovic, Z., Hou, Z., Biswas, K., & Muthukkumarasamy, V. (2024). Robust integration of blockchain and explainable federated learning for automated credit scoring. *Computer Networks*, 243, 110303.,

[30] Xu, Y., Kou, G., Peng, Y., Ding, K., Ergu, D., & Alotaibi, F. S. (2024). Profit- and risk-driven credit scoring under parameter uncertainty: A multiobjective approach. Omega, 125, 103004.,

[31] Hurlin, C., Pérignon, C., & Saurin, S. (2024). The fairness of credit scoring models. Management Science. Advance online publication.,

[32] Addy, W. A., et al. (2024). AI in credit scoring: A comprehensive review of models and predictive analytics. Global Journal of Engineering and Technology Advances, 18(2), 118–129., # Duplicate for completeness

[33] Rofik, R., et al. (2024). The optimization of credit scoring model using stacking ensemble learning and oversampling techniques. Journal of Information System Exploration and Research, 2(1).,

[34] Pertiwi, D. A. A., et al. (2024). Using genetic algorithm feature selection to optimize XGBoost performance in Australian credit. *Journal of Soft Computing Exploration*, 5(1), 92–98.

[35] Olola, T. M., & Olatunde, T. I. (2025). Artificial Intelligence in Financial and Supply Chain Optimization: Predictive Analytics for Business Growth and Market Stability in The USA. International Journal of Applied Sciences and Radiation Research, 2(1). https://doi.org/10.22399/ijasrar.18

[36] R. Vidhya, D. Lognathan, S, S., P.N. Periyasamy, & S. Sumathi. (2025). Anomaly Detection in IoT Networks Using Federated Machine Learning Approaches. *International Journal of Computational and Experimental Science and Engineering,* 11(3). https://doi.org/10.22399/ijcesen.2485

[37] Fowowe, O. O., & Agboluaje, R. (2025). Leveraging Predictive Analytics for Customer Churn: A Cross-Industry Approach in the US Market. *International Journal of Applied Sciences and Radiation Research*, 2(1). https://doi.org/10.22399/ijasrar.20

[38] Makin , Y., & Pavan K Gondhi. (2025). A Quantitative Framework for Portfolio Governance Using Machine Learning Techniques. *International Journal of Computational and Experimental Science and Engineering,* 11(3). https://doi.org/10.22399/ijcesen.2474

[39] Hafez, I. Y., & El-Mageed, A. A. A. (2025). Enhancing Digital Finance Security: AI-Based Approaches for Credit Card and Cryptocurrency Fraud Detection. *International Journal of Applied Sciences and Radiation Research*, 2(1). https://doi.org/10.22399/ijasrar.21

[40] Shanmugam Muthu, R S, N., A. Tamilarasi, Ahmed Mudassar Ali, S, S., & S. Jayapoorani. (2025). AI-Powered Predictive Digital Twin Platforms for Secure Software-Defined IoT Networks. *International Journal of Computational and Experimental Science and Engineering*, 11(3). https://doi.org/10.22399/ijcesen.2497

[41] Ibeh, C. V., & Adegbola, A. (2025). AI and Machine Learning for Sustainable Energy: Predictive Modelling, Optimization and Socioeconomic Impact In The USA. *International Journal of Applied Sciences and Radiation Research*, 2(1). https://doi.org/10.22399/ijasrar.19

[42] Sharad Kelkar , A. (2025). Study and Analysis of AI and Fintech on Quality of Accounting Information Disclosures and Corporate Governance with special reference to Banking Sector. *International Journal of Computational and Experimental Science and Engineering*, 11(2). https://doi.org/10.22399/ijcesen.2229