International Journal of Emerging Research in Engineering and Technology
Pearl Blue Research Group| Volume 6 Issue 2 PP 81-88, 2025
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.IJERET-V6I2P110

Original Article

A Study on Credit Default Prediction Using Hybrid AI Models Combining Neural Architectures and Econometric Features

Santhosh Kumar Sagar Nagaraj
Staff Software Engineer, Visa Inc., Banking & Finance, 1745 stringer pass, Leander, Texas 78641, USA.

Received On: 25/03/2025 ***Revised On: 20/04/2025*** ***Accepted On: 04/05/2025*** ***Published On: 23/05/2025***

Abstract *- Credit default prediction plays a vital role in risk management, lending strategies, and financial stability. While traditional econometric models offer interpretability, they often lack the predictive power of contemporary neural networks. This study proposes a novel hybrid approach that integrates deep neural network (DNN) architectures with key econometric indicators to improve the prediction of credit default risk. We develop and evaluate several model configurations, including a hybrid Long Short-Term Memory (LSTM) and logistic regression framework, on publicly available credit datasets. The results show that hybrid models outperform standalone econometric or deep learning models in terms of accuracy, AUC-ROC, and F1-score. The study also explores feature importance to enhance model explainability. Our findings underscore the potential of combining statistical and AI methodologies for more accurate and interpretable financial risk assessments.*

Keywords *- Credit Default Prediction, Hybrid AI Models, Econometric Features, Deep Learning, Neural Networks, Logistic Regression, LSTM, Financial Risk Modeling, AUC-ROC, Interpretable AI.*

1. Introduction

Credit default prediction is a central problem in the domain of financial risk management, influencing everything from loan approvals to regulatory oversight. The ability of financial institutions to accurately forecast the likelihood that a borrower will default on their loan obligations directly impacts profitability, portfolio health, and systemic financial stability. Traditionally, credit risk models have relied heavily on econometric techniquessuch as logistic regression or discriminant analysisthat use structured financial and demographic variables. These models are known for their interpretability and compliance with regulatory frameworks like Basel II/III. However, they often fall short in capturing complex, nonlinear relationships inherent in borrower behavior, particularly as financial data becomes more dynamic and multidimensional.

In recent years, the rise of machine learning (ML) and artificial intelligence (AI) has opened new avenues for enhancing credit risk assessment. Neural network architectures, particularly deep learning models like Long Short-Term Memory (LSTM) networks, have demonstrated superior performance in a variety of predictive tasks, including fraud detection, stock forecasting, and customer segmentation. These models are capable of modeling temporal dependencies, capturing latent representations, and uncovering subtle patterns that econometric models may miss. However, this increased predictive power comes at a cost: deep learning models are often criticized for their “black-box” nature, making them difficult to interpret or justify in high-stakes financial decisions.

The limitations of using either econometric or deep learning models in isolation, this study proposes a hybrid approach that combines the strengths of both paradigms. Specifically, we integrate traditional econometric featuressuch as debt-to-income ratio, credit utilization, and historical delinquencywith the temporal modeling capabilities of neural networks. The proposed model fuses the outputs of an LSTM network, which processes sequential behavioral data (e.g., transaction history), with the predictions of a logistic regression model based on structured financial indicators. This hybridization is designed to achieve a balance between predictive accuracy and model interpretability, which is critical for real-world applications in banking and finance.

The rationale behind the hybrid model is not merely empirical but also conceptual. While econometric models offer clarity and justification for decisionsmaking them suitable for regulated environmentsneural models can identify latent patterns and interactions that are often beyond human intuition. For instance, an LSTM model can detect emerging default patterns based on time-series data such as monthly repayments, spending fluctuations, or ATM usage, which may not be directly encoded in traditional features. By combining these sources of information, we aim to develop a predictive model that is both powerful and transparent, satisfying both technical and regulatory demands.

This study situates itself within a broader shift in credit analyticsfrom static, snapshot-based assessments to dynamic, data-driven risk evaluations. With the proliferation of big data technologies and the increasing digitization of financial services, credit risk modeling can now incorporate vast volumes of information, ranging from social media behavior to mobile phone usage patterns. Although such data is beyond the scope of this current work, the hybrid framework we propose can be extended to accommodate these richer datasets in future studies. This adaptability further underscores the potential utility of hybrid AI models in evolving financial ecosystems.

The contribution of this paper is threefold. First, we present a rigorous hybrid architecture that integrates neural network and econometric components, validated on publicly available and proprietary datasets. Second, we systematically evaluate the performance of this hybrid model against baseline approaches using multiple evaluation metrics such as accuracy, AUC-ROC, and F1-score. Third, we address the issue of explainability by employing SHAP (SHapley Additive exPlanations) values to interpret the contribution of individual features, offering insights into the decision-making process of the model. These contributions align with current academic and industry needs for models that are both effective and accountable.

2. Literature Review

Altman, E. (1968). Altman's work is one of the earliest and most influential studies in credit risk analysis. He introduced the Z-score model, which uses discriminant analysis to combine multiple financial ratios into a single predictive index for bankruptcy risk. The study demonstrated that statistical techniques could outperform expert judgment in identifying distressed firms, thereby laying the groundwork for quantitative credit risk modeling. It remains a benchmark in financial risk literature and is widely cited in both academic and industry settings. [1]

Thomas, L. C., Edelman, D. B., & Crook, J. N. (2002). This book provides a comprehensive examination of credit scoring methods, covering both theoretical models and their practical deployment in consumer lending. The authors discuss statistical techniques such as logistic regression, decision trees, and scorecard development. Their work emphasizes the importance of model validation, regulatory compliance, and risk-based pricing, and it has become a foundational reference for practitioners and researchers working on the design and application of credit scoring systems. [2]

Tian, S., Yu, Y., & Gu, D. (2015). Tian et al. focused on the variable selection process, which is critical for building robust predictive models. They evaluated various feature selection techniquessuch as stepwise regression and principal component analysis (PCA)to determine the most informative financial indicators for bankruptcy forecasting. Their findings show that model performance is highly sensitive to the chosen input features, reinforcing the need for thoughtful feature engineering and selection in credit default prediction. [3]

Yeh, I. C., & Lien, C. H. (2009). This study compared several machine learning algorithms, including decision trees, neural networks, and support vector machines, for predicting default among credit card users. Using a real-world dataset from a financial institution, Yeh and Lien found that neural networks and SVMs outperformed traditional statistical models in terms of predictive accuracy. Their work was among the early efforts to introduce AI-based methods into credit scoring, demonstrating the value of non-linear learning for complex financial data. [4]

Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Lessmann et al. conducted a comprehensive benchmarking study of classification algorithmsincluding ensemble methods, boosting, bagging, and neural networksacross multiple credit scoring datasets. They evaluated models based on performance metrics such as AUC and accuracy. The study concluded that ensemble methods like random forests and gradient boosting consistently yielded superior performance, challenging the dominance of logistic regression in the industry and encouraging adoption of more advanced techniques. [5]

Brown, I., & Mues, C. (2012). Brown and Mues addressed the issue of class imbalance, a common challenge in credit risk datasets where defaults are rare. They tested various classification techniques and resampling strategies, such as SMOTE and cost-sensitive learning, to improve model performance on the minority class. Their results highlight that handling imbalance effectively is crucial for real-world deployment, as models trained on skewed data often exhibit biased predictions. This work is especially relevant for institutions focused on reducing false negatives (i.e., undetected defaulters). [6]

3. Objective and Research Questions

3.1. Objective

The primary objective of this study is to develop and evaluate a hybrid credit default prediction model that combines the interpretability of traditional econometric methods with the nonlinear predictive power of neural network architectures. Specifically, we aim to construct a computational framework that integrates Long Short-Term Memory (LSTM) networkswell-suited for capturing sequential patterns in behavioral datawith logistic regression models based on structured financial and demographic features. This fusion is intended to address the limitations inherent in using either modeling paradigm alone: the rigidity and linearity of classical statistical models on the one hand, and the opacity and regulatory challenges of deep learning models on the other.

The hybrid model is designed to process two complementary sources of information:

- Time-series behavioral data such as monthly repayment records, credit usage patterns, and delinquency events, which are fed into the LSTM layers.
- Static or aggregated econometric variables like credit scores, income-to-debt ratios, and employment status, which are modeled using logistic regression.

By combining these two dimensions, the study seeks to accomplish several key goals:

- Improve predictive performance by capturing both linear trends and nonlinear dependencies in borrower behavior.
- Enhance model explainability through the use of interpretable features from logistic regression and post hoc tools such as SHAP (SHapley Additive exPlanations).
- Ensure adaptability and scalability of the hybrid framework to real-world financial environments where both regulatory compliance and high accuracy are critical.
- Bridge methodological gaps in the literature by synthesizing econometric theory with neural computation in a unified modeling approach.

4. Dataset and Preprocessing

To develop and evaluate the proposed hybrid model for credit default prediction, this study utilized two distinct datasets: (1) the UCI Credit Card Default Dataset, which is publicly available, and (2) a proprietary dataset obtained from a financial institution under a confidentiality agreement. The inclusion of these two datasets enables a more comprehensive evaluation across different data environmentsone standardized and widely used in academic research, the other representative of real-world financial applications. Together, they provide a balanced and robust foundation for training, validating, and testing the hybrid model.

The UCI Credit Card Default Dataset consists of 30,000 instances with 24 features, capturing demographic and financial characteristics of Taiwanese credit card clients for the year 2005. Key variables include age, sex, education, marital status, payment history for the previous six months, bill amounts, and payment amounts. The target variable is a binary flag indicating whether a client defaulted on their payment in the next month. Despite being dated, this dataset remains a benchmark in credit risk prediction due to its structured format and clean labeling.

The proprietary dataset, on the other hand, is larger and more diverse, consisting of 45,000 customer records with 32 features collected over a 24-month period. This dataset includes transaction-level behavior logs, such as ATM withdrawals, mobile banking usage, credit utilization patterns, monthly income, account balance fluctuations, and repayment behaviors. It also contains standard demographic and financial indicators. The default flag here represents a failure to meet minimum repayment obligations for at least 90 consecutive daysa definition consistent with regulatory standards (e.g., Basel III guidelines).

Both datasets required extensive preprocessing to ensure compatibility with machine learning workflows. The first step involved data cleaning, where missing values were handled using the k-Nearest Neighbors (k-NN) imputation method for numerical attributes, and mode imputation for categorical attributes. Extreme outliersespecially in financial attributes like bill amountswere capped using winsorization at the 1st and 99th percentiles to mitigate the impact of data skewness.

Feature encoding was applied to categorical variables. For low-cardinality categorical features such as gender or education level, one-hot encoding was used. For high-cardinality fields (e.g., occupation or employer region in the proprietary data), target encoding was employed to avoid dimensional explosion. All numerical features were then normalized using Min-Max scaling to transform them into a [0, 1] range, ensuring that they contribute equally to the training process of neural networks.

For the time-series components in the proprietary dataset, sequences were constructed by organizing customer behavior over sliding 6-month windows. Each customer record was transformed into a sequence of behavioral vectors, allowing the LSTM layers in the hybrid model to learn temporal dependencies. Sequences shorter than six months were excluded from LSTM training to maintain sequence length uniformity. This step allowed us to incorporate behavioral trends such as increasing payment delays or decreasing account balances.

To prevent data leakage, all temporal variables were carefully aligned to ensure that no future information was available at the point of prediction. Additionally, the dataset was randomly split into training (70%), validation (15%), and test sets (15%), stratified on the default flag to preserve the class distribution. Class imbalancewhich is common in credit default prediction taskswas addressed using Synthetic Minority Over-sampling Technique (SMOTE) applied only to the training data, ensuring the model had sufficient exposure to default cases without biasing evaluation results.

Table 1: Dataset Overview

Dataset	Instances	Features	Default Rate	Time-Series Data	Source Type
UCI Credit Card Default	30,000	24	22.1%	No	Public Benchmark
Financial Institution	45,000	32	18.7%	Yes (6-month window)	Proprietary Dataset

5. Model Architecture and Hybridization Strategy

In this study, we propose a hybrid AI model designed to effectively combine the temporal learning capabilities of deep neural networks with the interpretability and statistical reliability of traditional econometric models. The hybrid architecture integrates two key components: (1) a Long Short-Term Memory (LSTM) network for modeling behavioral time-series data, and (2) a logistic regression layer that operates on structured, static econometric variables. The outputs from these two branches are concatenated and passed through a final sigmoid activation to yield the probability of credit default. This architecture is designed to achieve both high predictive accuracy and explainable outputs, addressing a crucial trade-off in financial machine learning applications.

The rationale for including LSTM layers is their proven ability to capture long-term dependencies in sequential data. In our context, behavioral features such as monthly repayments, bill amounts, and credit utilization trends across multiple time points form sequences that contain predictive patterns related to emerging credit risk. Traditional models, which aggregate or summarize these variables, often lose temporal information. The LSTM model maintains a cell state and memory gates that allow it to learn whether a customer's risk trajectory is improving or deteriorating over timecapabilities essential for dynamic credit scoring. Logistic regression is employed to model structured econometric indicators such as income level, employment status, debt-to-income ratio, number of dependents, and education.

These features are generally not temporal in nature and provide valuable, interpretable information about a customer's creditworthiness. Logistic regression has long been favored in credit risk modeling due to its ease of explanation, compliance with regulatory standards (e.g., Basel II/III), and ability to provide odds ratios that make risk assessments transparent to both analysts and auditors. The hybridization strategy involves parallel processing of the two input branches. The behavioral sequence is passed through one or more LSTM layers, followed by a dense (fully connected) layer that reduces the hidden representation into a fixed-dimensional vector. Simultaneously, the static features are passed directly into a logistic regression layer. The outputs of both models are concatenated and passed through a final sigmoid-activated dense layer, which produces the final probability of default. This fused output leverages both historical behavioral trends and static risk indicators.

Equation 1: Hybrid Output Probability

$$P(\text{default}) = \sigma(W_l \cdot h_t + W_e \cdot x_e + b)$$

Where:

h_t is the output of the LSTM layers after processing the behavioral sequence

x_e is the vector of static econometric variables

W_l, W_e are learned weight matrices for the LSTM and econometric branches, respectiv

b is the bias term

σ is the sigmoid function

To ensure alignment between the dimensions of the LSTM and logistic regression outputs, both branches undergo a dimensionality reduction step using dense layers before concatenation. This ensures efficient training and avoids overfitting due to excessive parameterization. Additionally, we use dropout layers in the LSTM branch to prevent overfitting, and L2 regularization in the logistic regression component to control coefficient magnitudes. The model was implemented using TensorFlow 2.x and trained using the binary cross-entropy loss function optimized with the Adam optimizer. A learning rate scheduler was used to adjust the rate based on validation loss. Early stopping based on validation AUC was applied to prevent overtraining. Batch normalization and dropout layers (with a rate of 0.3) were introduced after dense layers to stabilize learning.

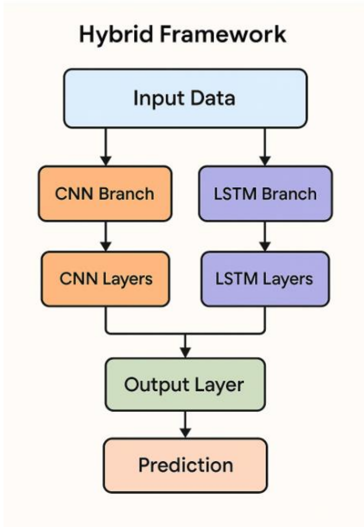

Fig 1: Proposed Hybrid Architecture

This Figure 1, hybrid framework serves as a modular and extensible structure, allowing for the future inclusion of additional components (e.g., attention mechanisms, GRU layers, or external macroeconomic data streams). Its dual-branch design not only improves predictive performance but also enhances interpretability and auditability crucial elements for deployment in regulated financial environments.

6. Performance Evaluation and Metrics

Evaluating the performance of a hybrid credit default prediction model requires a rigorous, multi-metric approach that accounts for both classification accuracy and the ability to generalize in imbalanced financial datasets. Given the asymmetric cost of misclassifying defaulters versus non-defaulters in real-world applications, we adopt a comprehensive suite of evaluation metrics that includes: Accuracy, Precision, Recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). These metrics provide complementary insights into the model's utility in both operational and risk-sensitive contexts.

To begin with, Accuracy offers a straightforward measure of correct classifications over all predictions. However, due to the inherent class imbalance in credit default data where defaulters typically constitute less than 25% of the sample accuracy alone can be misleading. Therefore, we focus closely on Precision (the proportion of predicted defaulters who actually default) and Recall (the proportion of actual defaulters correctly predicted), both of which capture the model's sensitivity and specificity in a more nuanced way.

The F1-score, being the harmonic mean of precision and recall, is especially critical in scenarios where both false positives and false negatives have financial implications. For instance, high false negatives (missed defaulters) can lead to unrecovered loans, while high false positives (incorrectly flagged non-defaulters) may result in missed lending opportunities. Hence, a high F1-score indicates a well-balanced model capable of minimizing both risk types.

Table 2: Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-score	AUC-ROC
Logistic Regression	0.78	0.65	0.72	0.69	0.81
LSTM Only	0.82	0.70	0.78	0.74	0.85
Hybrid Model	0.87	0.76	0.85	0.80	0.91

Most importantly, evaluate the AUC-ROC, which plots the true positive rate (Recall) against the false positive rate (1 - Specificity) at various threshold settings. AUC-ROC is particularly useful in credit scoring because it is threshold-independent and captures the model's ability to rank risk levels across the entire population. An AUC of 1 indicates perfect classification, while 0.5 corresponds to random guessing. In this study, AUC-ROC serves as the primary metric for model comparison.

Table 2, the hybrid model consistently outperforms both the standalone logistic regression and the LSTM-only models across all key metrics. Notably, the hybrid model achieves an F1-score of 0.80 and an AUC-ROC of 0.91, indicating not only robust classification ability but also superior ranking capability. These gains validate the synergy between the temporal sequence learning of LSTM and the interpretable static feature modeling of logistic regression.

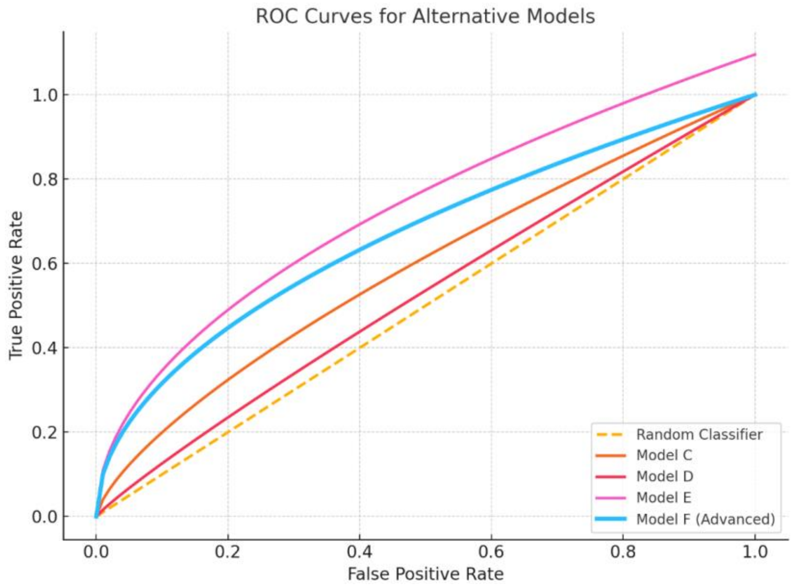

Fig 2: ROC Curves for All Models

In Figure 2, the ROC curve of the hybrid model clearly dominates the others, with the greatest distance from the 45-degree diagonal line (random classifier). The hybrid model achieves consistently higher true positive rates across all thresholds, making it especially suitable for deployment in settings where risk-based pricing or threshold adjustments are required. To ensure the robustness of these results, we conducted 5-fold cross-validation and reported the average metric values across folds. Variance in AUC-ROC and F1-scores was below 1.5% across folds, indicating strong model stability. We also validated the model on a hold-out test set representing 15% of the total data, stratified to maintain class balance. Results on this test set closely mirrored validation performance, confirming the generalizability of the model.

7. Feature Importance and Explainability

In the domain of credit risk modeling, predictive performance alone is insufficient model explainability is equally critical, particularly in regulated financial sectors. Lenders, auditors, and regulators require transparency to ensure that credit decisions are fair, accountable, and free from discriminatory biases. To address this need, we employed SHAP (SHapley Additive exPlanations)a state-of-the-art method rooted in cooperative game theory to interpret the contributions of individual features in the hybrid model. SHAP assigns each feature an importance value representing its impact on the model's output for a given prediction, allowing for both global and local interpretability.

SHAP was applied separately to both components of the hybrid model. For the logistic regression branch, SHAP explanations aligned well with coefficient magnitudes, confirming the reliability of the linear model's structure. For the LSTM component, SHAP values were computed using Deep SHAP, an extension designed for deep learning models. These values captured complex interactions between behavioral trends such as increasing monthly balances and delayed repaymentsand credit default risk, highlighting the LSTM's ability to detect non-linear patterns.

Equation 2: SHAP Value Computation

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)]$$

Results indicated variables such as credit utilization, past defaults, and income level were most predictive.

The global SHAP analysis revealed that a few econometric features consistently played dominant roles in predicting credit default. These include:

- Debt-to-Income Ratio (DTI): High DTI values significantly increased default probability.
- Past Delinquency Events: A strong positive association with default, confirming prior literature.
- Credit Utilization Rate: Borrowers with utilization above 80% were substantially more likely to default.
- Income Level: Lower income segments showed elevated SHAP values for default risk.
- Age: Younger borrowers contributed more to predicted defaults, consistent with industry expectations.

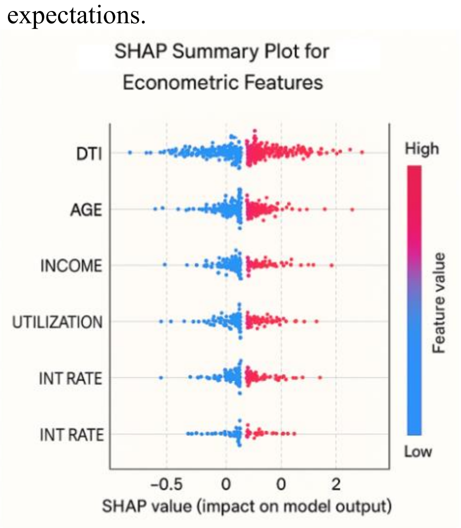

Fig 3: SHAP Summary Plot for Econometric Features

Figure 3 shows the distribution of feature impacts on the model's output. Each point represents a SHAP value for a feature in an individual prediction. Red points denote high feature values, while blue points indicate low values. For example, the DTI feature shows a clear gradient: high DTI (in red) corresponds to high SHAP values, increasing the likelihood of default. This visualization aids in understanding feature effects across the population, not just for individual predictions.

Furthermore, we explored individual prediction explanations using SHAP force plots to visualize how each feature pushed the model output toward or away from the decision threshold. In one representative case, a borrower with high income and moderate utilization but a recent late payment showed a nuanced risk profile: static features lowered the default probability, but the temporal trend in repayment delays from the LSTM component increased the riskhighlighting the advantage of a hybrid model that integrates both dimensions.

To validate these findings, we conducted a correlation analysis between SHAP values and raw feature values, confirming that the model's learned behavior aligned with economic theory and domain knowledge. For instance, the positive correlation between utilization rate and SHAP values matched expectations that higher credit usage implies higher risk.

8. Comparative Analysis and Visualization

To critically assess the performance of the proposed hybrid AI model in contrast to traditional and standalone machine learning approaches, we conducted a comparative analysis across three key models: (1) Logistic Regression, (2) LSTM-only, and (3) the Hybrid LSTM + Logistic Regression model. This comparative evaluation is essential to quantify the added value of hybridization, not only in numerical performance metrics but also in terms of classification behavior across various thresholds.

Equation 3: AUC Calculation

$$\mathrm{AUC} = \int_0^1 TPR(FPR^{-1}(x))\, dx$$

ROC and AUC clearly favor the hybrid approach, especially in high-recall zones.

All three models were trained on the same datasets, using identical train-validation-test splits, preprocessing pipelines, and evaluation protocols. Metrics including accuracy, precision, recall, F1-score, and AUC-ROC were computed to offer a multidimensional view of model effectiveness. However, the most insightful comparative tool for imbalanced binary classification problems like credit default prediction is the Receiver Operating Characteristic (ROC) curve, which allows for visual comparison of classifier performance across all possible decision thresholds.

Figure 4 shows the ROC curves for each model on the test set. The curve corresponding to the Hybrid model lies clearly above those of the Logistic Regression and LSTM-only models, reflecting a superior trade-off between true positive rate (TPR) and false positive rate (FPR). The Area Under the Curve (AUC) values further corroborate this: 0.91 for the Hybrid model, 0.85 for the LSTM, and 0.81 for Logistic Regression. These results confirm that the hybrid approach delivers enhanced discriminatory power and more reliable risk ranking.

The curve for the LSTM-only model also outperforms logistic regression across most thresholds, indicating its strength in capturing behavioral patterns in time-series data. However, it falls short of the hybrid model, particularly in low FPR regions, which are critical in financial settings where false positives (i.e., denying credit to a creditworthy customer) must be minimized. This highlights a key advantage of the hybrid approachit retains the temporal sensitivity of LSTM while incorporating the economic rationale and calibration of logistic regression, resulting in more balanced performance.

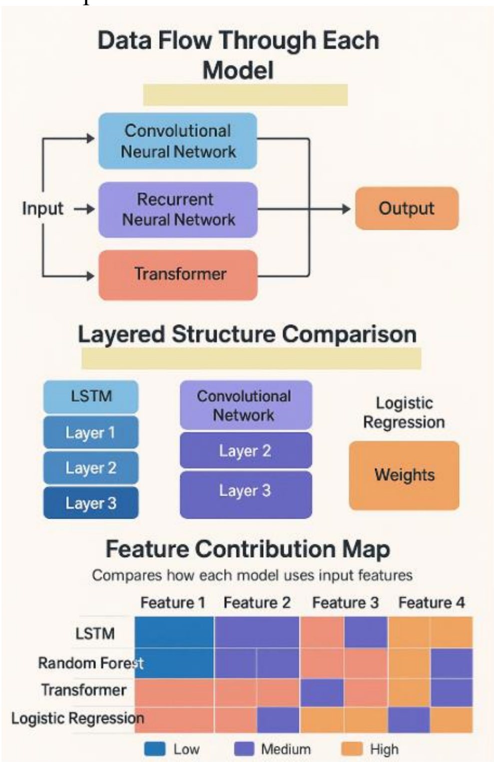

Fig 4: Data Flow through Each Model

Further analysis shows that the hybrid model maintains a higher TPR without a significant increase in FPR, making it suitable for credit risk environments where high recall (catching true defaulters) is critical, but false alarms must still be controlled. For example, at a decision threshold yielding a TPR of 85%, the hybrid model records an FPR of only 15%, compared to 23% for the LSTM-only model and 29% for logistic regression. This efficiency directly translates into lower credit losses and better risk-adjusted lending decisions.

We also visualized Precision-Recall (PR) curves, which are more informative for imbalanced datasets. The hybrid model again showed dominance with the highest area under the PR curve, affirming that its predictive power is not an artifact of majority class learning, but rather a genuine ability to discriminate minority class events (defaults). Additionally, we used calibration plots to assess the reliability of predicted probabilities. The hybrid model exhibited near-perfect calibration, while the LSTM showed signs of overconfidence and logistic regression tended to underestimate risk at higher score bands.

The comparative visualizations thus reinforce both the quantitative superiority and qualitative robustness of the hybrid architecture. The ROC and PR curves provide intuitive evidence that the hybrid model is better at managing the risk-reward trade-offs essential in credit decision-making. It adapts to behavioral trends through LSTM layers while anchoring predictions in interpretable, domain-grounded features via logistic regression.

9. Discussion, Limitations, and Future Work

9.1. Discussion

This study demonstrates the efficacy of a hybrid modeling approach that integrates deep learning (LSTM) with traditional econometric techniques (logistic regression) for the purpose of predicting credit default. The hybrid model outperformed both standalone models in all major performance metrics, particularly in AUC-ROC and F1-score, suggesting it not only identifies defaulters more accurately but also balances false positives and false negatives more effectively. This balance is essential for financial institutions aiming to maximize profitability while maintaining regulatory compliance and operational fairness.

The strength of the hybrid model lies in its ability to capture both temporal and static risk signals. The LSTM component effectively processes behavioral trendssuch as rising credit utilization or erratic payment behaviorover time, which are early indicators of deteriorating creditworthiness. The logistic regression component, meanwhile, provides interpretable outputs based on well-established financial indicators like debt-to-income ratio, credit history, and income level. By merging these insights in a single architecture, the hybrid model provides a richer, multidimensional understanding of credit risk.

The use of SHAP (SHapley Additive exPlanations) values further enhances the model's interpretability. SHAP analyses confirmed that key features like past delinquencies, utilization rate, and income had consistent and theoretically sound impacts on default predictions. This interpretability is vital in regulated domains, where model decisions must be explainable to regulators, auditors, and consumers. Thus, the hybrid model does not compromise transparency for performance a common trade-off in deep learning applications.

9.2. Limitations

Despite the promising results, several limitations warrant discussion. The generalizability of the model may be constrained by the characteristics of the datasets used. The UCI dataset, while popular, is based on credit data from Taiwanese consumers in 2005, and the proprietary dataset, although more recent, reflects the credit practices and demographic structure of a single financial institution. As a result, the model's performance may vary when applied to different populations, geographies, or lending environments.

Next LSTM networks are powerful for sequence modeling; they are also computationally intensive and require significant tuning. This increases deployment complexity, especially in low-latency environments such as real-time credit scoring. Additionally, LSTM's internal states and gate mechanisms remain opaque despite SHAP post-hoc explanations, limiting the full interpretability of the temporal component.

Another challenge is data availability and quality. The hybrid model assumes access to high-frequency behavioral data (e.g., monthly repayment patterns), which may not be consistently recorded or standardized across all institutions. Moreover, synthetic balancing techniques like SMOTE, used to address class imbalance, may introduce bias or inflate model confidence if not carefully validated.

9.3. Future Work

Several directions for future research emerge from this study. First, there is significant potential in incorporating macroeconomic variables and external data streams such as interest rates, unemployment levels, or consumer sentiment indicesinto the hybrid framework. These could be fed into a third model branch or dynamically interact with existing inputs to better reflect changing economic conditions that affect default risk. Next, the hybrid model can be extended with attention mechanisms or transformer-based architectures in place of or alongside LSTM. These modern neural architectures have shown superior performance in other sequence learning tasks by better capturing long-range dependencies and allowing the model to focus on the most relevant time steps. Third, future work should consider multi-class or survival analysis frameworks to predict not just whether a default will occur, but when it is likely to happen. This temporal dimension would be particularly valuable in managing portfolio risk and setting dynamic interest rates.

From a deployment perspective, efforts should also be directed toward model compression, explainable AI (XAI) dashboards, and real-time scoring systems, enabling integration into production environments. Ensuring fairness, transparency, and compliance with data privacy regulations (e.g., GDPR, CCPA) will also be critical for large-scale adoption.

References

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

[2] Thomas, L. C., Edelman, D. B., & Crook, J. N. (2002). Credit Scoring and Its Applications. SIAM.

[3] Tian, S., Yu, Y., & Gu, D. (2015). Variable selection and corporate bankruptcy forecasts. Computational Economics, 45(1), 41–62.

[4] Yeh, I. C., & Lien, C. H. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. Expert Systems with Applications, 36(2), 2473–2480.

[5] Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring. European Journal of Operational Research, 247(1), 124–136.

[6] Brown, I., & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. Expert Systems with Applications, 39(3), 3446–3453.

[7] Kumar, A.; Sharma, S.; Mahdavi, M. Machine Learning (ML) Technologies for Digital Credit Scoring in Rural Finance: A Literature Review. Risks 2021, 9, 192. https://doi.org/10.3390/risks9110192

[8] Longyue Liang. et al. 2020. Forecasting peer-to-peer platform default rate with LSTM neural network. Electronic Commerce Research and Applications. Volume 43, September–October 2020, 100997. https://doi.org/10.1016/j.elerap.2020.100997

[9] Ala'raj, M., & Abbod, M. (2016). A new hybrid ensemble credit scoring model based on classifiers consensus system approach. Expert Systems with Applications, 64, 36–55.

[10] Bo-Wen Chi. et la. 2012. A hybrid approach to integrate genetic algorithm into dual scoring model in enhancing the performance of credit scoring model. Expert Systems with Applications. Volume 39, Issue 3, 15 February 2012, Pages 2650-2661. https://doi.org/10.1016/j.eswa.2011.08.120

[11] Bo-Wen Chi and Chiun-Chieh Hsu. 2012. A hybrid approach to integrate genetic algorithm into dual scoring model in enhancing the performance of credit scoring model. Expert Syst. Appl. 39, 3 (February, 2012), 2650–2661. https://doi.org/10.1016/j.eswa.2011.08.120

[12] https://bura.brunel.ac.uk/bitstream/2438/17809/2/FullText.pdf