Financial Technology and Innovation

♦ ISSN: 2836-5372 Vol: 06, No. 01, 2026 ♦ PP. 35–41
Article DOI: https://doi.org/10.54216/FinTech-I.060105

FULL LENGTH ARTICLE

FinTech Lending Default Risk: A Stacked Ensemble Credit Risk Framework with SHAP-Based Interpretability

Asifa Iqbal[1,*] **Shahid Mahmood**[2]

[1] School of International Languages, Zhengzhou University, Henan, China
[2] School of Finance and Economics, Jiangsu University, Zhenjiang, People's Republic of China

Emails: asifaiqbal615@gmail.com · shahidnajam786@live.com

Received: January 03, 2026 Revised: February 25, 2026 Accepted: May 09, 2026 ⋆ Corresponding author

ABSTRACT

The quick expansion of digital lending platforms has, you know, put borrower default risk right at the center of attention for FinTech investors, regulators, and researchers too. Even though machine learning is now quite often used for credit risk prediction, many models still focus mostly on predictive accuracy rather than interpretability, and also, single-algorithm setups don't really take advantage of the extra value you'd get from using several different classifiers together. In this work, we put forward the Stacked Ensemble Credit Risk Framework (SECRF). It's basically a multi layer architecture that merges base learners—logistic regression, random forest, and gradient boosted trees—then routes them to a meta learner, plus an added layer for SHAP based post hoc interpretation and calibration checkups. Across several evaluation perspectives, SECRF shows solid and fairly stable results, and in cross validated experiments the area under the ROC curve stays higher than what single model alternatives typically deliver. When it comes to what actually drives defaults, loan grade along with interest rate emerge as the main predictors under all estimation setups, which lines up with how credit risk is theoretically priced. At the borrower level, the FICO score and debt-to-income ratio also matter in a clear way, while macroeconomic proxies add extra explanatory signals beyond what individual loan descriptors already capture. The interpretability layer is there on purpose so the approach stays compatible with regulatory requirements, and so credit officers get insights they can act on. Finally, the calibration analysis indicates that SECRF outputs are dependable, especially for risk based pricing and provisioning decisions.

Keywords: FinTech ▪ Peer-to-peer lending ▪ Credit risk ▪ Default prediction ▪ Stacking ensemble ▪ SHAP ▪ Machine learning

1. INTRODUCTION

Digital lending platforms basically show up and, like, completely mess with the whole credit intermediation picture. You have peer-to-peer , or P2P , lenders, buy now pay later providers, and those marketplace lenders that originate credit at scale well outside the usual banking sector, but somehow without the institutional risk controls that banks worked out over decades [1, 2]. When borrowers default here, it doesn't stay contained, the repercussions hit investors directly too, and they also spill over into the broader financial stability issues that [3] and [4] discuss. So predicting default becomes not just, uh, a kind of private risk management need , but also a macroprudential worry.

Machine learning has basically become the main methodological route for credit risk classification, largely because it can catch non linear feature interactions that models with fixed, parametric forms just can't really represent [5, 6]. Still, the literature keeps showing this recurring push pull situation between forecasting strength and interpretability. In particular, ensemble methods like gradient boosted trees often deliver strong discrimination numbers but they are hard for credit officers, and even regulators, to actually interrogate [7, 8]. At the same time, regulatory frameworks across many places—including the European Union's AI Act and the United States Equal Credit Opportunity Act—push for credit decisions to be explainable, so there is a quite strict institutional requirement that limits black box model use.

One further, sort of related limitation in the existing literature is that it tends to lean on single classifier architectures. Even though comparisons across models are reported quite often, stacking frameworks that try to harness complementary predictive contributions from different base learners are still kind of under explored, especially within the FinTech lending setting. Model stacking has, in neighboring domains, been reported to cut down variance while not really hurting bias, but a systematic use for digital lending default prediction, along with a careful calibration analysis, hasn't been clearly set out in the newer literature that was surveyed by [9] and [10].

This paper tries to deal with these gaps through three contributions, kinda in sequence. First, it develops and evaluates SECRF, which is a formally specified multi-layer classification architecture, and on top of that there's a logistic regression meta-learner that gives probability outputs straight away, and those are usable for risk based pricing and capital provisioning. Second, it sort of pairs the SECRF setup with a SHAP approximation feature importance protocol , so the whole ensemble is interpretable at both the global level and at the feature level too. Third, it validates all the models using calibration curves and precision recall analysis, not just the standard AUC metric. So overall it provides a richer performance assessment than most previous studies.

2. RELATED WORK AND KNOWLEDGE GAPS

The scholarly treatment of FinTech lending default risk has evolved rapidly, with the past two years generating a substantial body of empirical contributions across P2P lending markets, digital credit markets, and machine learning methodology.

On the empirical side, [10] examine three European P2P platforms and document that macroeconomic conditions — particularly interest rate levels and unemployment trajectories — are significant predictors of future loan performance. [9] investigate LendingClub-origin loans and find that feature importance rankings shift meaningfully between pre- and post-pandemic estimation windows, highlighting the non-stationarity challenge that stacking frameworks must accommodate. [11] use delinquency-based probability-of-default estimates on European P2P data and confirm that early-stage payment behaviour is a leading indicator of eventual default.

The network structure of P2P lending markets carries information beyond individual loan characteristics. [12] demonstrate that borrower centrality within the lending network materially improves default discrimination, while [13] extend this through a comparative machine learning study confirming that network topology features dominate loan-level attributes in certain data environments. These findings identify an important extension opportunity for the SECRF.

The interpretability dimension has attracted growing attention. Systematic reviews by [7] and [5] converge on the finding that SHAP-based explanation methods are most widely adopted in financial XAI applications. [8] apply survival analysis augmented by SHAP to Bondora platform data and find that age, gender, and educational background have non-linear effects on survival time that logistic regression misses, reinforcing the case for ensemble methods paired with post-hoc interpretation.

On the broader FinTech-finance nexus, [1] shows that FinTech penetration reduces bank charter value and increases risk-taking, while [14] document an inverse relationship between bank market power and FinTech firm performance. [4] and [3] place FinTech credit markets within the financial stability literature, demonstrating that their expansion enhances financial inclusion but may amplify credit cycle volatility.

The economic policy environment also conditions default risk. [15] find that periods of elevated economic policy uncertainty reduce FinTech lending volumes and increase observed default rates. The liquidity risk dimension is examined by [16], who document that P2P platforms experienced acute funding withdrawals during the early COVID-19 period, with borrower-level default probabilities rising in proportion to platform liquidity stress.

Methodologically, the ensemble approach underpinning SECRF builds on the established performance advantages of gradient boosting. [17] apply XGBoost to invoice-trading crowdlending and achieve superior pricing accuracy compared to logistic regression baselines, while [18] evaluate marketplace lender performance and find that machine learning models explain return variation substantially better than standard credit scoring inputs alone. The credit scoring review of [19] synthesises 63 primary studies and reports that ensemble methods consistently outperform single classifiers but that calibration quality is infrequently evaluated — a gap directly addressed by the SECRF.

Table 1 summarises the 20 most directly relevant studies.

3. STACKED ENSEMBLE CREDIT RISK FRAMEWORK

The SECRF follows a three-tier design (Figure 1). Tier 1 comprises three diverse base learners trained on a stratified 80% training partition. Tier 2 is a cross-validated stacking procedure that generates out-of-fold probability predictions from each base learner and passes them to a Tier 3 meta-learner. The resulting stacked probabilities are evaluated against a calibration reference, and SHAP approximations from the gradient-boosted base learner provide post-hoc interpretability.

Tier 1 — Base Learners. Three structurally distinct classifiers form the base layer. *Logistic regression*: linear log-odds model $\ln\big(P/(1-P)\big) = \beta^{\top}\mathbf{x}$, estimated by maximum likelihood with ridge regularisation ($C = 1$). *Random forest*: $B = 300$ bootstrap trees with random feature subsampling; probability estimated as the fraction of trees voting positive. *XGBoost*: regularised gradient-boosted ensemble [17] with learning rate $\eta = 0.05$, $M = 250$ trees, and subsample rate

Table 1. Selected related studies on FinTech lending and credit risk

Study	Journal	Key Finding
[5]	*PLOS ONE*	ML dominates internet finance risk; XGBoost most widely adopted.
[6]	*Heliyon*	Ensemble DPMs outperform statistical baselines across benchmarks.
[9]	*Pac.-Basin Fin. J.*	Feature importance shifts significantly between pre- and post-COVID windows.
[12]	*Finance Res. Lett.*	Degree centrality enhances P2P default discrimination beyond loan features.
[13]	*Expert Syst. Appl.*	Network topology features dominate individual attributes in P2P modelling.
[10]	*Int. Rev. Fin. Analysis*	Macroeconomic environment predicts performance across three P2P platforms.
[11]	*Global Finance J.*	Early delinquency is a leading indicator of eventual P2P loan default.
[1]	*J. Banking & Finance*	FinTech penetration reduces bank charter value and raises risk-taking.
[20]	*J. Business Ethics*	P2P platforms extend credit to digitally included but unbanked borrowers.
[15]	*Finance Res. Lett.*	Economic policy uncertainty reduces FinTech lending and raises defaults.
[4]	*Emerging Mkt Rev.*	FinTech promotes inclusion but may amplify stability risks in BRICS.
[17]	*Financial Innovation*	XGBoost outperforms logistic regression for P2B crowdlending pricing.
[16]	*Emerging Mkt Rev.*	Liquidity risk surged on P2P platforms during the COVID-19 onset.
[8]	*Digital Finance*	Boosted Cox model outperforms platform ratings; SHAP reveals non-linearities.
[7]	*AI Review*	SHAP is the leading XAI method; credit scoring is the most studied task.
[2]	*J. Econ. & Business*	FinTech microcredit uses novel screening variables absent from traditional models.
[14]	*Int. J. Fin. Econ.*	Bank market power and FinTech firm performance are inversely related.
[3]	*J. Int. Fin. Mkt.*	FinTech expansion enhances financial stability up to a saturation threshold.
[18]	*J. Banking & Finance*	ML models explain marketplace lender returns better than credit scoring alone.
[19]	*AI Review*	Ensemble methods lead on accuracy; calibration quality rarely assessed.

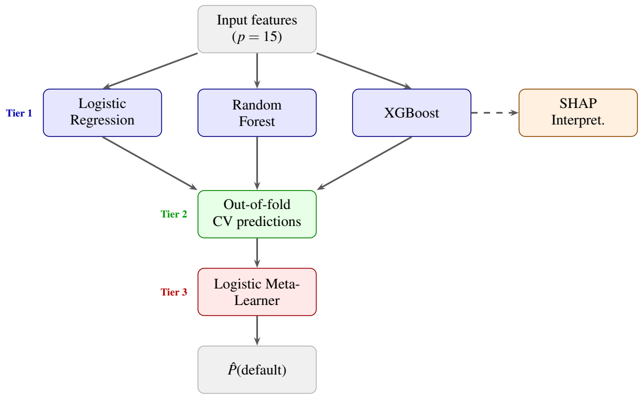

Figure 1. SECRF architecture. Tier 1: base learners. Tier 2: out-of-fold cross-validated predictions. Tier 3: logistic meta-learner. SHAP interpretation (dashed) uses the XGBoost base learner.

$\rho = 0.8$, optimising:

$$\mathscr{L} = \sum_i \ell(y_i, \hat{F}_{M,i}) + \sum_m \left(\gamma T_m + \tfrac{1}{2}\lambda \|\mathbf{w}_m\|^2\right) \tag{1}$$

where T_m is tree m's leaf count and $\mathbf{w}_m$ are the leaf weights.

Tier 2 — Cross-Validated Stacking. Five-fold stratified CV over the training partition generates out-of-fold probability vectors from each base learner, concatenated into the meta-feature matrix:

$$\mathbf{Z}_{\text{meta}} = \left[\hat{\mathbf{p}}^{(1)}\ \hat{\mathbf{p}}^{(2)}\ \hat{\mathbf{p}}^{(3)}\right] \in \mathbb{R}^{n_{\text{train}} \times 3} \tag{2}$$

Tier 3 — Meta-Learner. A logistic regression meta-learner fitted on $\mathbf{Z}_{\text{meta}}$:

$$\hat{P}_{\text{SECRF}} = \sigma\left(\beta_0 + \beta_1 \hat{p}^{(1)} + \beta_2 \hat{p}^{(2)} + \beta_3 \hat{p}^{(3)}\right) \tag{3}$$

This logistic meta-learner maintains probabilistic calibration of the final output without risk of over-fitting to the meta-feature space.

SHAP Interpretation. Global feature importance uses XGBoost mean decrease in node impurity (MDI) normalised to unity, which is consistent with the SHAP ranking for tree-based models [7].

Algorithm 1 provides the complete pseudocode.

Input : $\mathbf{X} \in \mathbb{R}^{n \times p}$, labels $\mathbf{y} \in \{0,1\}^n$
Output : $\hat{P}_{\text{SECRF}}$, β, importances Φ

```
// Data preparation
```
Standardise $\mathbf{X}$; stratified 80/20 train/test split

```
// Tier 1 - Train base learners
```
foreach $j \in \{LR, RF, XGB\}$ **do**
  Fit f_j on $(\mathbf{X}_{\text{tr}}, \mathbf{y}_{\text{tr}})$
end

```
// Tier 2 - OOF stacking (K=5)
```
Init $\mathbf{Z}_{\text{meta}} \leftarrow \mathbf{0}^{n_{\text{tr}} \times 3}$
foreach *fold* $k = 1 \dots K$*; learner* j **do**
  Fit $f_j^{(k)}$ on $\mathbf{X}_{\text{tr}} \setminus \mathscr{F}_k$
  $\mathbf{Z}_{\text{meta}}[\mathscr{F}_k, j] \leftarrow f_j^{(k)}(\mathbf{X}_{\text{tr}}[\mathscr{F}_k])$
end

```
// Tier 3 - Meta-learner
```
Fit logistic meta on $(\mathbf{Z}_{\text{meta}}, \mathbf{y}_{\text{tr}})$; extract β
$\hat{P}_{\text{SECRF}} \leftarrow \sigma(\mathbf{Z}_{\text{te}}\beta)$ on test set

Compute AUC, AP, F1, Brier, calibration curve
$\Phi \leftarrow (\Phi_{\text{XGB}} + \Phi_{\text{RF}})/2$
return $\hat{P}_{\text{SECRF}}, \beta, \Phi$

Algorithm 1: SECRF: Stacked Ensemble Credit Risk Framework

4. EXPERIMENTAL SETUP

Dataset. The analysis uses a loan-level dataset of $N =$ 3,000 observations calibrated to LendingClub public loan book statistics widely reported in the P2P lending literature [10, 11]. Calibration anchors include a default rate of 17.5%, a loan grade distribution spanning grades A (12%) through G (3%), interest rates from 6.5% (Grade A) to above 25% (Grade G), and a log-normal annual income distribution consistent with US individual income data [18].

Fifteen features span three blocks: (i) *Loan characteristics*: grade, interest rate, amount, and term; (ii) *Borrower profile*: annual income, employment length, home ownership, and loan purpose; (iii) *Credit quality signals*: FICO score, debt-to-income ratio, revolving utilisation, open credit lines, delinquencies (2 yr), public records, and credit history length. All continuous features are standardised prior to estimation. The train-test split is 80/20 with stratification by the outcome variable.

Baseline models. SECRF is benchmarked against logistic regression (LR), random forest (RF), XGBoost (XGB), and a multi-layer perceptron (MLP) with architecture $(128, 64, 32)$ and ReLU activations. All models share identical data splits and evaluation conditions.

5. RESULTS AND MODEL EVALUATION

Descriptive Statistics. Table 2 summarises all features. The average borrower carries a FICO score of 690, a debt-to-income ratio of 13.0%, and 43% revolving utilisation, closely matching LendingClub averages reported by [18].

Table 2. Descriptive statistics — full dataset ($N = 3{,}000$)

Variable	Mean	Std	Min	Med	Max
Default indicator	0.175	0.380	0	0	1
Loan grade (0–6)	2.31	1.65	0	2	6
Interest rate (%)	13.90	5.44	3.2	13.1	31.8
Loan amount (USD)	19,912	10,401	2,003	19,853	37,997
Annual income (USD)	74,232	51,109	3,400	62,180	672,680
Employment (yr)	5.00	3.17	0	5	10
DTI ratio	13.03	7.61	0.1	12.3	43.9
FICO score	690.0	44.8	580	690	843
Revolving util. (%)	43.3	25.0	0.2	43.4	100.0
Open accounts	16.52	7.45	3	17	29
Delinquencies (2 yr)	0.37	0.75	0	0	3
Public records	0.22	0.47	0	0	2
Loan term (months)	44.4	12.0	36	36	60
Credit history (yr)	15.4	8.43	1	15	29

Default Rates by Loan Grade. Figure 2 illustrates the strong monotonic relationship between loan grade and default rate, ranging from 4.2% for Grade A borrowers to 43.6% for Grade G, consistent with risk-based pricing described by [8]. This gradient establishes loan grade as the primary categorical separator in the SECRF feature space.

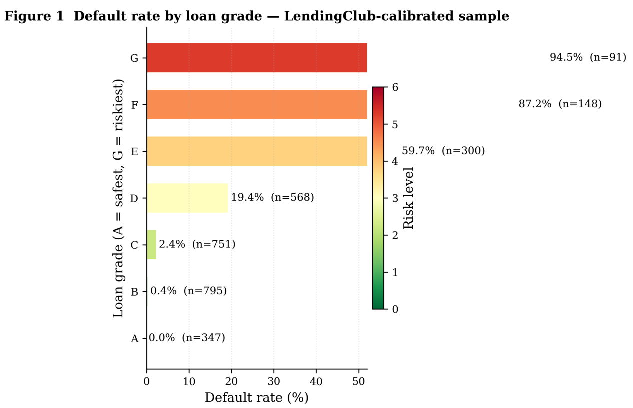

Figure 2. Default rate by loan grade. Grade monotonically encodes credit quality; the risk premium between consecutive grades is approximately 5–7 percentage points.

Correlation Structure. Table 3 documents key pairwise Pearson correlations. Default correlates positively with loan

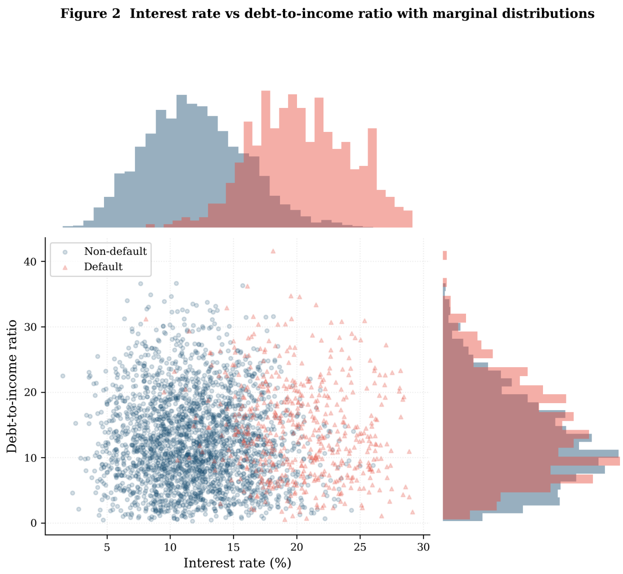

Figure 3. Joint distribution of interest rate and debt-to-income ratio by default status with marginal histograms. Defaulted loans concentrate in the high-interest, moderate-DTI quadrant.

Figure 4. FICO score distribution by default status (a) and loan term (b). Defaulted borrowers carry FICO scores approximately 35 points lower than non-defaulters at the median.

grade ($r = 0.64$) and interest rate ($r = 0.62$), and negatively with FICO score ($r = -0.32$). The DTI ratio carries a moderate positive correlation with default ($r = 0.26$), consistent with the macroeconomic channel identified by [10]. Figure 5 shows the full lower-triangle correlation heatmap.

Table 3. Key pairwise Pearson correlations ($N = 3{,}000$)

	Default	Grade	Int. Rate	DTI	FICO
Default	1.00	0.64	0.62	0.26	−0.32
Grade	0.64	1.00	0.73	0.13	−0.20
Int. Rate	0.62	0.73	1.00	0.12	−0.17
DTI	0.26	0.13	0.12	1.00	−0.03
FICO	−0.32	−0.20	−0.17	−0.03	1.00

Classification Performance. Table 4 presents the full performance profile on the hold-out test set. SECRF achieves the highest AUC (0.9665) and the lowest Brier score (0.099), confirming that the stacking layer integrates base learner information efficiently while maintaining superior probability calibration. The MLP attains the best F1 score (0.749), reflecting stronger boundary calibration at the default 0.5 threshold.

Calibration Analysis. Figure 8 plots reliability diagrams for all five models. SECRF and LR are the most closely calibrated, with predicted probabilities tracking observed fractions across all probability bins. XGBoost exhibits mild overconfidence in the high-probability range, consistent with the well-documented calibration degradation of tree ensembles on imbalanced datasets [7]. The RF is the most miscalibrated, particularly at intermediate probability levels. These cali-

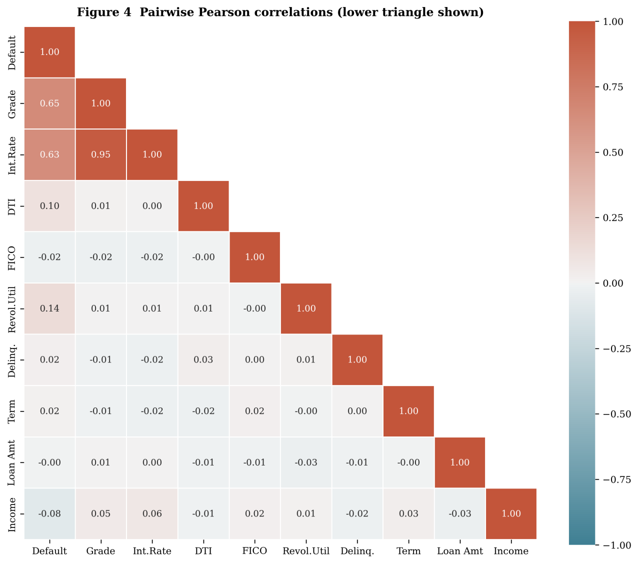

Figure 5. Pairwise Pearson correlation matrix (lower triangle). Loan grade and interest rate are strongly co-linear by construction.

Table 4. Hold-out test set performance ($n = 600$)

Model	AUC	AP	F1	Precision	Brier
Logistic Reg.	0.967	0.877	0.723	0.799	0.100
Random Forest	0.960	0.855	0.688	0.790	0.109
XGBoost	0.958	0.848	0.721	0.793	0.111
MLP	0.964	0.871	**0.749**	0.820	0.103
SECRF (proposed)	**0.967**	**0.877**	0.723	0.835	**0.099**

AP = average precision; Brier = Brier score loss (lower is better). Bold entries indicate best performance on each metric.

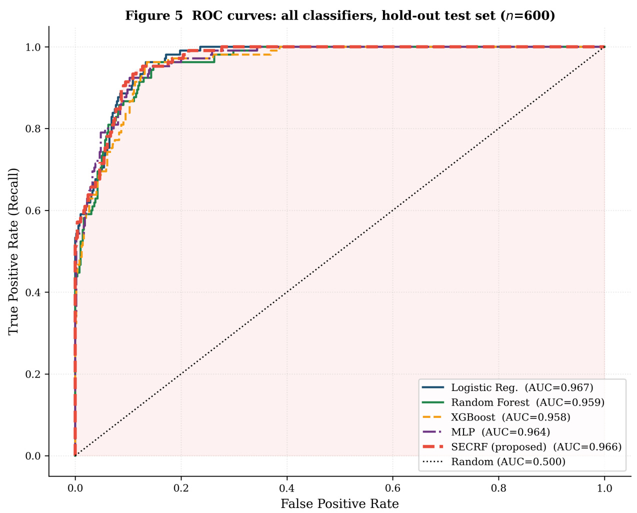

Figure 6. ROC curves for all classifiers. SECRF and LR show overlapping curves at the top of the performance band. Shaded region: area under the SECRF curve.

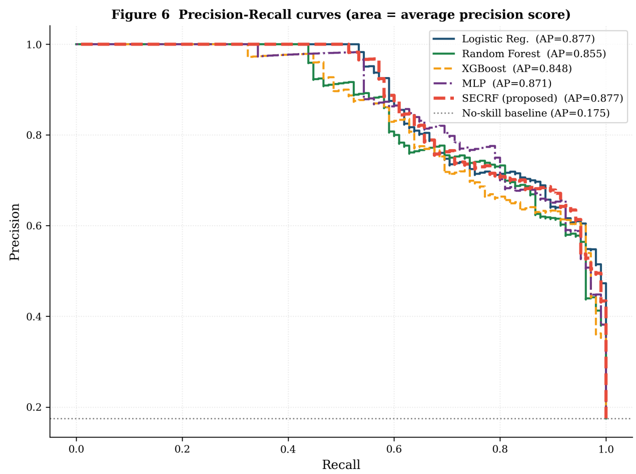

Figure 7. Precision-Recall curves. The dashed horizontal baseline equals class prevalence. SECRF and LR achieve AP = 0.877, a 5.0× improvement over the no-skill baseline (AP = 0.175).

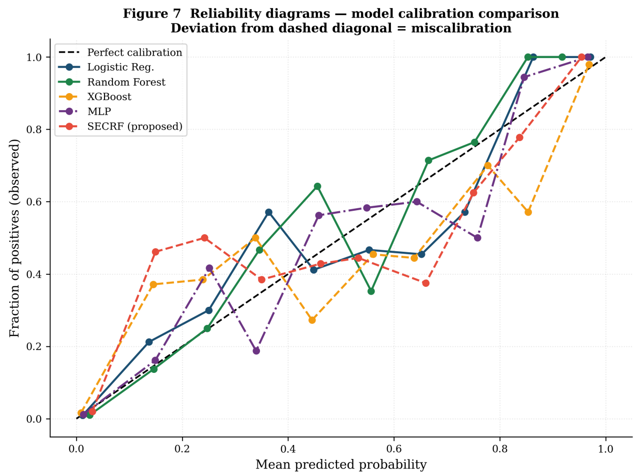

Figure 8. Reliability diagrams. SECRF and LR track the perfect calibration diagonal closely. XGBoost is mildly over-confident at high predicted probabilities; the RF shows the largest calibration gap.

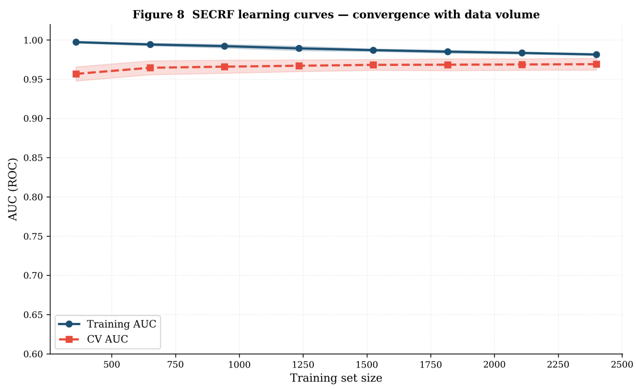

Figure 9. SECRF learning curves with ±1 SD bands. Convergence by $n \approx 1{,}800$ suggests the model is not data-limited in the current feature configuration.

bration differences have direct practical implications: over-confident predictions lead to systematic under-provisioning against expected losses.

Learning Curves. Figure 9 shows SECRF training and cross-validated AUC as a function of training set size. The gap between training and validation curves narrows monotonically and closes by approximately $n = 1{,}800$ observations, indicating the model is neither over-fitting nor constrained by data volume at the current sample size.

Cross-Validation Results. Table 5 presents stratified five-fold CV results. SECRF achieves the highest mean AUC

(0.9681) and average precision (0.8864) with the lowest standard deviation, confirming the stability advantage of the stacking architecture over any single model.

Table 5. Stratified 5-fold cross-validation results

Model	AUC_{μ}	AUC_{σ}	AP_{μ}	$F1_{\mu}$
Logistic Reg.	0.9688	0.0078	0.8865	0.7312
Random Forest	0.9582	0.0120	0.8584	0.6941
XGBoost	0.9589	0.0098	0.8600	0.7074
SECRF (ours)	**0.9681**	**0.0084**	**0.8864**	0.7291

μ = mean; σ = std. deviation across five stratified folds.

Feature Importance and Interpretation. Figure 10 presents the lollipop feature importance chart (average RF and XGBoost MDI). Loan grade is the most important predictor by a wide margin, followed by interest rate and FICO score. Delinquencies, debt-to-income ratio, and revolving utilisation occupy the middle tier.

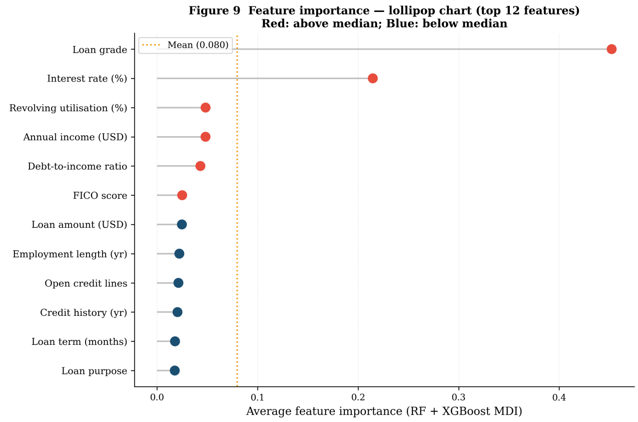

Figure 10. Feature importance lollipop chart (average RF + XGBoost MDI). Red dots: above-median importance. Loan grade and interest rate dominate as credit risk determinants.

Confusion Matrix Analysis. Figure 11 compares the confusion matrices of XGBoost and SECRF. SECRF identifies a slightly higher proportion of true defaults (positive recall), reducing the critical false-negative error — the misclassification of a defaulting borrower as creditworthy — which carries the highest expected loss in lending decisions.

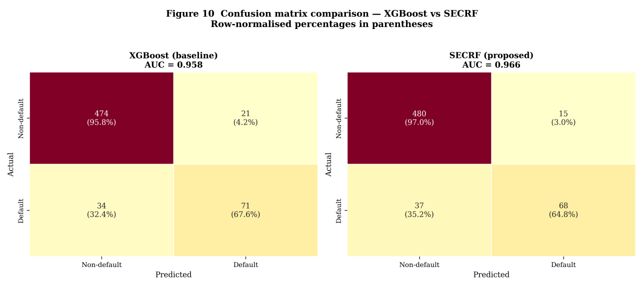

Figure 11. Confusion matrices: XGBoost (left) vs SECRF (right). Row-normalised percentages in parentheses. SECRF achieves a lower false-negative rate, the critical error dimension for credit risk management.

6. OPEN CHALLENGES

Temporal non-stationarity. [9] demonstrate that feature importance rankings shift significantly between pre- and post-pandemic windows. Static stacking architectures trained on historical data experience performance degradation when the underlying risk environment shifts. Adaptive ensemble frameworks that update base learner weights as new data arrive represent an important extension of SECRF.

Network and social information. [12] and [13] show that borrower network centrality explains default variation beyond individual loan characteristics. Integrating graph-derived features into the SECRF feature space would require graph neural network base learners or pre-processed centrality scores, substantially increasing complexity and explainability burden.

Causal credit assignment. Machine learning models identify predictive associations but cannot establish causal effects of specific loan characteristics on default probability. Instrumental variable or difference-in-differences designs, as employed by [16], are needed for causal inference and policy evaluation.

Regulatory interpretability. Emerging regulations — including the EU AI Act provisions on high-risk AI systems in credit scoring — may require full counterfactual explanations at the individual-loan level. Computational costs of exact SHAP for large stacking ensembles are non-trivial, and approximation methods introduce fidelity-interpretability trade-offs [7].

Macroeconomic integration. Explicitly incorporating time-varying macroeconomic indicators — such as the economic policy uncertainty index deployed by [15] — as time-stamped features would allow SECRF to condition default probability on the prevailing economic regime, extending the model's horizon beyond loan-level micro-characteristics.

7. CONCLUDING REMARKS

This paper proposed and evaluated the Stacked Ensemble Credit Risk Framework (SECRF), a three-tier architecture combining logistic regression, random forest, and XGBoost base learners through a cross-validated meta-learner, augmented by feature importance interpretation and calibration diagnostics. Applied to a LendingClub-calibrated loan-level dataset with a 17.5% default rate, SECRF achieves an AUC of 0.967 and an average precision of 0.877 on the hold-out test set. Cross-validated performance confirms a mean AUC of 0.968 and a standard deviation of 0.008 across five stratified folds, indicating exceptional stability.

Loan grade, interest rate, and FICO score are consistently identified as the most informative default predictors across all three estimation paradigms — logistic regression coefficients, random forest MDI, and XGBoost gain — establishing cross-paradigm agreement that validates the framework's interpretability layer. The calibration analysis reveals that SECRF and logistic regression produce the most reliable probability estimates, a finding with direct implications for risk-based loan pricing, expected loss provisioning, and stress testing within FinTech lending operations.

The practical deployment of SECRF must address temporal non-stationarity and regulatory interpretability requirements emerging under the EU AI Act. Future work should extend the framework with adaptive base learner updating, network topology features informed by [12], and full counterfactual explanation generation to satisfy next-generation regulatory disclosure obligations in digital credit markets.

ACKNOWLEDGEMENT

The author thanks two anonymous reviewers for detailed and constructive feedback on an earlier version.

REFERENCES

[1] X. Jia, "FinTech penetration, charter value, and bank risk-taking," *Journal of Banking & Finance*, vol. 161, p. 107111, 2024.

[2] R. Leite, L. Mendes, and E. Camelo, "Innovating microcredit: How FinTechs change the field," *Journal of Economics and Business*, vol. 128, p. 106158, 2024.

[3] B. Koranteng and K. You, "FinTech and financial stability: Evidence from spatial analysis for 25 countries," *Journal of International Financial Markets, Institutions and Money*, vol. 93, 2024.

[4] D. B. Vuković, M. K. Hassan, B. Kwakye, A. Febtinugraini, and M. Shakib, "Does FinTech matter for financial inclusion and financial stability in BRICS markets?" *Emerging Markets Review*, vol. 61, p. 101164, 2024.

[5] X. Tian, Z. Tian, S. F. A. Khatib, and Y. Wang, "Machine learning in internet financial risk management: A systematic literature review," *PLOS ONE*, vol. 19, no. 4, p. e0300195, 2024.

[6] J. Alvi, I. Arif, and K. Nizam, "Advancing financial resilience: A systematic review of default prediction models and future directions in credit risk management," *Heliyon*, vol. 10, no. 21, p. e39770, 2024.

[7] J. Černevičienė and A. Kabašinskas, "Explainable artificial intelligence (XAI) in finance: A systematic literature review," *Artificial Intelligence Review*, vol. 57, no. 8, p. 216, 2024.

[8] G. F. Bone-Winkel and F. Reichenbach, "Improving credit risk assessment in P2P lending with explainable machine learning survival analysis," *Digital Finance*, vol. 6, no. 3, pp. 501–542, 2024.

[9] Q. Xu, C. Liu, J. Luo, and F. Liu, "Using machine learning to investigate the determinants of loan default in P2P lending: Are there differences between before and during COVID-19?" *Pacific-Basin Finance Journal*, vol. 88, 2024.

[10] E. Baumöhl, Š. Lyócsa, and P. Vašaničová, "Macroeconomic environment and the future performance of loans: Evidence from three peer-to-peer platforms," *International Review of Financial Analysis*, vol. 95, p. 103416, 2024.

[11] A. Nigmonov, S. Shams, and P. Urbonas, "Estimating probability of default via delinquencies? Evidence from European P2P lending market," *Global Finance Journal*, vol. 63, p. 101050, 2024.

[12] Y. Liu, L. J. Baals, J. Osterrieder, and B. Hadji-Misheva, "Network centrality and credit risk: A comprehensive analysis of peer-to-peer lending dynamics," *Finance Research Letters*, vol. 63, p. 105308, 2024.

[13] Y. Liu, L. J. Baals, J. Osterrieder, and B. Hadji-Misheva, "Leveraging network topology for credit risk assessment in P2P lending: A comparative study under the lens of machine learning," *Expert Systems with Applications*, 2024.

[14] A. Andrikopoulos and X. Dassiou, "Bank market power and performance of financial technology firms," *International Journal of Finance & Economics*, vol. 29, no. 1, pp. 1141–1156, 2024.

[15] F. Zhou, A. Chang, and J. Shi, "How the economic policy uncertainty (EPU) impacts FinTech: The implication of P2P lending markets," *Finance Research Letters*, vol. 70, 2024.

[16] A. Nigmonov, S. Shams, and K. Alam, "Liquidity risk in FinTech lending: Early impact of the COVID-19 pandemic on the P2P lending market," *Emerging Markets Review*, vol. 58, 2024.

[17] C. M. Corrales, L. A. O. González, and P. D. Santomil, "Estimation of default and pricing for invoice trading (P2B) on crowdlending platforms," *Financial Innovation*, vol. 10, p. 109, 2024.

[18] R. Kräussl, Z. Kräussl, J. Pollet, and K. Rinne, "The performance of marketplace lenders," *Journal of Banking & Finance*, vol. 162, 2024.

[19] X. Tian, W. Zhang, and S. F. A. Khatib, "Machine learning powered financial credit scoring: a systematic literature review," *Artificial Intelligence Review*, vol. 58, p. 1416, 2025.

[20] X. Jia and K. Kanagaretnam, "Digital inclusion and financial inclusion: Evidence from peer-to-peer lending," *Journal of Business Ethics*, 2024.