INTERNATIONAL JOURNAL OF ACTA INFORMATICA
VOLUME (2024)

EMBEDDING ADAPTIVE MACHINE-LEARNING DECISION SUPPORT IN SUSPICIOUS ACTIVITY REPORT INVESTIGATION WORKFLOWS: AN ENTERPRISE INFORMATION-SYSTEMS FRAMEWORK

Md Soebur Rahman[1*], Eun-Ye Kim[2]
[1] International American University, USA
[2]Ewha Womans University, SOUTH KOREA

*Corresponding Author Email: soebrahman10@gmail.com

ABSTRACT

Community Development Financial Institutions (CDFIs) now represent a $452 billion industry of 1,487 Treasury-certified institutions as of the first quarter of 2023, having grown rapidly, industry assets nearly tripled and the certified-institution count rose 40 percent, over the preceding five years, driven substantially by depository institutions, especially credit unions, newly pursuing CDFI certification, even as small-business loan demand and CDFI reliance both grew alongside this expansion. This article proposes a decision-support architecture for embedding early-risk-signaling analytics into CDFI credit-risk and loan-monitoring workflows for risk-aware small-business lending, grounding the proposal in published, cited machine-learning results for small-business credit-risk prediction, including neural-network models reported to achieve 95 percent accuracy and a 0.98 ROC-AUC compared with 79 percent accuracy and a 0.58 AUC for logistic regression on comparable small-business credit data, and in real, current market context showing three-quarters of CDFIs reporting increased demand for their products in the most recent full-industry survey. The article reviews CDFI industry structure and its small-business lending role, details the proposed architecture and its grounding in published early-warning and credit-risk literature, and provides a candid assessment of what remains unvalidated specifically for CDFI-scale institutions and portfolios

KEYWORDS: Data Quality; Data Governance; FAIR Principles; Data Lifecycle; Artificial Intelligence Ethics; Healthcare Data; Data Management; ISO Standards

INTRODUCTION

Community Development Financial Institutions occupy a distinctive and, per recent Federal Reserve research, rapidly growing position within the U.S. small-business lending landscape. Treasury's CDFI Fund, established in 1994, has certified 1,487 CDFIs as of the first quarter of 2023, together holding 452 billion U.S. dollars in assets as of the same period [1]. An August 2023 Federal Reserve Bank of New York report documents that both figures have grown substantially over the preceding five years: CDFI industry assets nearly tripled between 2018 and 2023, and the number of certified institutions rose 40 percent over the same window, growth the report attributes substantially to more depository institutions, credit unions and banks, newly pursuing CDFI certification. CDFI-certified credit unions in particular nearly doubled in number, from 290 in 2019 to 529 in 2023, and CDFI credit unions remain the largest single segment of the industry by assets, holding roughly 300 billion U.S. dollars, or about 66 percent of total industry assets, with banks holding 118 billion U.S. dollars (27 percent) and loan funds holding 35 billion U.S. dollars (8 percent) [1].

This rapid growth in CDFI capacity has occurred alongside, and reinforced by, evidence of growing demand for exactly the mission-driven small-business lending CDFIs specialise in. The Federal Reserve's 2023 CDFI Survey, fielded April through June 2023 and gathering responses from 453 CDFIs, nearly a third of the industry, found three-quarters of respondent CDFIs reported increased demand for their products over the prior year, with a similar share expecting demand to continue increasing into the following year [2]. Among CDFIs that saw increased or steady demand, 40 percent reported fully meeting it, 42 percent mostly, and 18 percent only partially, indicating that demand growth is already testing many CDFIs' capacity to serve it fully [2]. Eighty-four percent of survey respondents reported total assets of 375 million U.S. dollars or less, underscoring how small the typical CDFI remains even amid the sector's overall growth [2].

This combination, a rapidly growing CDFI industry, sustained small-business demand for CDFI capital, and CDFIs' persistently small typical institution size, creates a specific and pressing operational challenge: CDFIs are being asked to serve more borrowers, often with thinner credit files and higher inherent risk than conventional bank borrowers, without a proportional increase in institutional capacity to do so. Embedding early-risk-signaling analytics into CDFI credit-risk and loan-monitoring workflows, allowing a CDFI's limited staff capacity to be deployed proportionately toward the borrowers most likely to need attention, is one concrete response to this pressure, and is the subject this article addresses.

This article proposes a decision-support architecture for this purpose, grounding the proposal specifically in published, cited machine-learning literature on small-business credit-risk prediction rather than illustrative projections. Section 2 reviews CDFI industry structure and its small-business lending role in greater depth. Section 3 presents the proposed decision-support architecture. Section 4 reviews published evidence on machine-learning approaches to small-business credit-risk and early-warning prediction. Section 5 discusses data and feature considerations specific to the CDFI context. Section 6 assesses the evidence base candidly. Section 7 addresses governance and adoption considerations, and Section 8 concludes.

WHY CDFI UNDERWRITING DIFFERS FROM CONVENTIONAL BANK LENDING

Before turning to methodology, it is worth being explicit about why CDFI credit-risk analytics warrant distinct treatment from the mainstream community-bank early-warning literature this article series addresses elsewhere. CDFIs are certified specifically because they serve borrowers, low-income individuals, minority-owned businesses, and enterprises in economically distressed communities, that conventional lenders systematically decline, generally due to thin credit files, limited collateral, or a risk profile that standardised underwriting models penalise disproportionately relative to the borrower's actual repayment capacity. This mission-driven adverse selection is not a flaw in the CDFI model; it is the model's entire purpose. But it means that any credit-risk analytics architecture built for CDFIs must be evaluated against a borrower population that is, by design, riskier and thinner-filed than the general small-business population the studies reviewed in Section 4 were developed against, a distinction Section 6 returns to directly when assessing the evidence base's limitations.

METHODOLOGY AND SOURCE SELECTION

Sources cited in this article were identified through targeted searches for primary Federal Reserve research on CDFI industry structure and small-business lending conditions, industry-analyst reporting on CDFI financial performance (specifically Aeris Insight's annual CDFI sector analysis), and peer-reviewed and preprint academic literature on machine-learning approaches to small-business and SME credit-risk and default prediction. Given that no source identified in this research evaluates a machine-learning early-risk model specifically on CDFI-originated loan data, this article draws its technical evidence from the broader small-business and SME credit-risk literature and states this limitation explicitly in Section 6, rather than implying CDFI-specific validation where none exists.

CDFI INDUSTRY STRUCTURE AND SMALL-BUSINESS LENDING ROLE

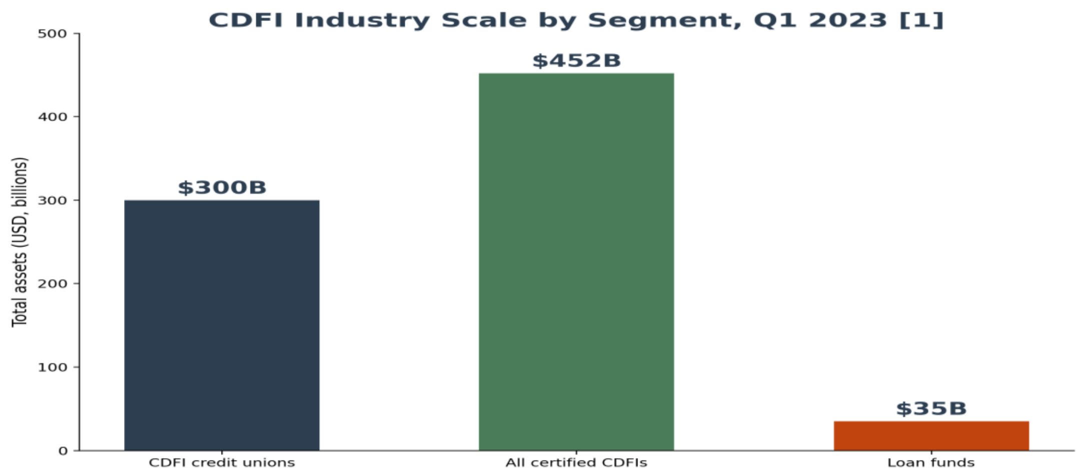


Figure 1 presents the CDFI industry's scale by segment as of Q1 2023. CDFI credit unions, holding roughly 300 billion U.S. dollars, represent the largest asset base within the industry, though these institutions are primarily focused on consumer financial services rather than small-business lending specifically [1]. Certified loan funds, the segment most directly associated with small-business and microenterprise lending, represented 39 percent of certified institutions by count (582 of 1,487) as of Q1 2023, still the largest single category by count even as the credit-union segment's share of total assets has grown fastest over the preceding five years [1].

The Federal Reserve's 2023 CDFI Survey provides further useful structural context: loan funds continue to represent around half of survey respondents, and 23 percent of CDFIs identifying as loan funds reported total assets of 5 million U.S. dollars or less [2]. This scale, materially smaller than the mainstream community banks and credit unions discussed elsewhere in this article series, is a central design constraint for any analytics architecture proposed for this segment: an early-risk-signaling system for a typical small-business-focused CDFI loan fund must assume minimal dedicated data-science staffing, and correspondingly must prioritise ease of integration and interpretability over model sophistication for its own sake.

The same 2023 survey data further notes that, while CDFIs overall reported strong demand growth in 2023, meeting that demand fully remained a challenge for a meaningful share of institutions, only 40 percent of CDFIs with increased or steady demand reported fully meeting it [2], a gap the survey attributes partly to funding and capacity constraints rather than to lending appetite. This demand-capacity gap in the specific segment most relevant to this article, small-business lending, reinforces the case for efficiency-focused interventions, such as the early-risk-signaling architecture proposed here, that can help a resource-constrained small-business lending operation extend its limited capacity further, rather than assuming continued growth in staffing or funding will resolve the capacity pressure on its own.

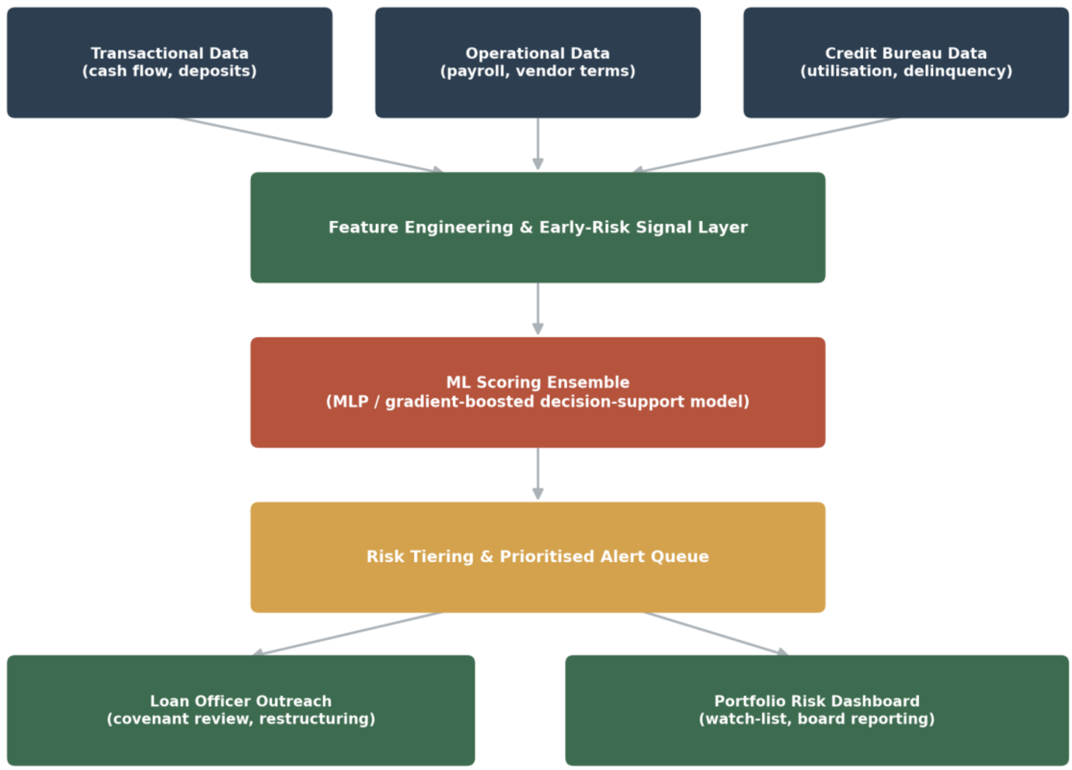
RURAL CDFIS AND THE SMALL-BUSINESS CONCENTRATION PATTERN

The Federal Reserve's 2023 survey work also surfaces a geographic pattern directly relevant to this article's architecture: rural CDFIs report small-business finance as a primary business line at a notably higher rate than nonrural CDFIs, while consumer financing represents a comparable share of primary business activity for both groups [2]. Rural CDFI respondents were also somewhat more likely than nonrural respondents to be smaller-asset institutions, and reported proximity to funders as a distinctly greater operational challenge than nonrural CDFIs did [2]. This geographic concentration pattern matters for the architecture proposed in Section 3: a disproportionate share of the small-business-focused CDFI segment this article addresses operates in rural markets with distinctly constrained access to the kind of philanthropic and public capital that might otherwise fund a dedicated internal data-science capability, reinforcing the case, developed further in Section 5, for an architecture that assumes minimal in-house technical capacity and prioritises data sources the institution can access at low marginal cost.

PROPOSED DECISION-SUPPORT ARCHITECTURE

Figure 2 presents the proposed architecture. Three data categories, transactional data reflecting cash flow and deposit activity, operational data reflecting payroll and vendor-payment behaviour, and credit-bureau data reflecting utilisation and delinquency, feed a feature-engineering and early-risk signal layer. This layer feeds a machine-learning scoring ensemble, for which Section 4 reviews published evidence supporting a multi-layer-perceptron or gradient-boosted architecture specifically, whose output feeds a risk-tiering and prioritised alert queue. This queue in turn routes to two complementary outputs: direct loan-officer outreach for covenant review and potential restructuring, and a portfolio-level risk dashboard supporting board and funder reporting.

Figure 2. Decision-Support Architecture for Embedding Early-Risk-Signaling Analytics into CDFI Credit-Risk Workflows



As with the architecture's predecessor version in this article series, the design deliberately routes every model output through a human loan officer or portfolio manager rather than permitting any automated lending or restructuring decision, reflecting both the qualitative, relationship-based judgement central to CDFI underwriting philosophy and the practical reality that a scoring model, however well validated on the broader literature reviewed in Section 4, has not itself been validated on CDFI-specific data, a limitation Section 6 addresses directly.

WHY THE ARCHITECTURE PRESERVES EXISTING RULE-BASED SCREENING

A further design principle evident in Figure 2, though not immediately visible in the diagram itself, is that the proposed architecture is intended to sit alongside, not replace, whatever rule-based or manual screening process a CDFI loan fund already operates. Given the evidence-strength limitations documented in Section 6, specifically the absence of any CDFI-specific validation for the models reviewed in Section 4, a CDFI adopting this architecture should initially treat its output as an additional, prioritising signal layered on top of existing underwriting and monitoring practice, rather than as a validated replacement for that practice. This mirrors the phased-adoption principle recommended throughout this article series: a new analytics layer earns expanded trust and expanded decision-making weight over time, as the institution accumulates its own, CDFI-specific evidence of the model's real-world reliability, rather than being granted that trust and weight from the moment of initial deployment based solely on the general, non-CDFI-specific literature reviewed in Section 4.

INTEGRATION WITH EXISTING CDFI REPORTING OBLIGATIONS

CDFIs already maintain substantial reporting infrastructure to satisfy CDFI Fund compliance and impact-reporting requirements, tracking metrics such as loans originated in target markets, jobs supported, and borrower demographic composition, obligations that exist independently of any credit-risk-analytics initiative. The architecture in Figure 2 is deliberately designed so that its portfolio-risk-dashboard output can draw on, and feed back into, this existing reporting infrastructure rather than requiring a wholly separate reporting pipeline, since building a duplicate reporting system solely for credit-risk analytics would impose exactly the kind of disproportionate technology burden Section 2's scale data suggests a typical small-business CDFI loan fund, nearly a quarter of which report total assets of 5 million U.S. dollars or less, is poorly positioned to absorb.

PUBLISHED EVIDENCE ON MACHINE-LEARNING APPROACHES TO SMALL-BUSINESS CREDIT RISK

A substantial and growing body of published research addresses machine-learning approaches to small-business and SME credit-risk and default prediction, though, as Section 6 discusses, none of the specific studies reviewed here evaluate performance on CDFI-originated loans specifically. Table 1 summarises key published results.

Table 1. Published Results for Machine-Learning Small-Business and SME Credit-Risk Prediction

Study / Source	Method & Data	Reported Result
Comparative small-business credit-risk analysis [5]	Comparison of logistic regression, random forest, XGBoost, and multi-layer-perceptron neural network models on small-business credit data	MLP: 95% accuracy, 0.98 ROC-AUC. Logistic regression: 79% accuracy, 0.58 ROC-AUC, 22% defaulter recall. Random forest: 68% defaulter recall. XGBoost: 86% accuracy, 0.74 ROC-AUC
WSMOTE-ensemble small-business loan default study [6]	Random-forest classifier combined with a weighted SMOTE sampling technique addressing severe class	15.16% improvement in minority (default) class accuracy versus competing methods; sampling

Study / Source	Method & Data	Reported Result
	imbalance in small-business default data	methods outperformed non-sampling approaches
Agricultural Bank of Egypt loan-default study [8]	Random forest, decision tree, and gradient-boosting models on customer loan data, with embedded and recursive feature selection	Decision tree: 88% accuracy, 53% recall, 89% specificity. Loan balance, due amount, and delinquency history identified as most predictive features
Interpretable ML for Italian SME credit default [7]	Review and application of interpretable machine-learning methods to SME default prediction, building on established early-warning literature	Cites foundational early-warning literature (Laitinen & Chong, 1999; Lin, Ansell & Andreeva, 2012) establishing financial-distress-definition sensitivity in small-business default prediction

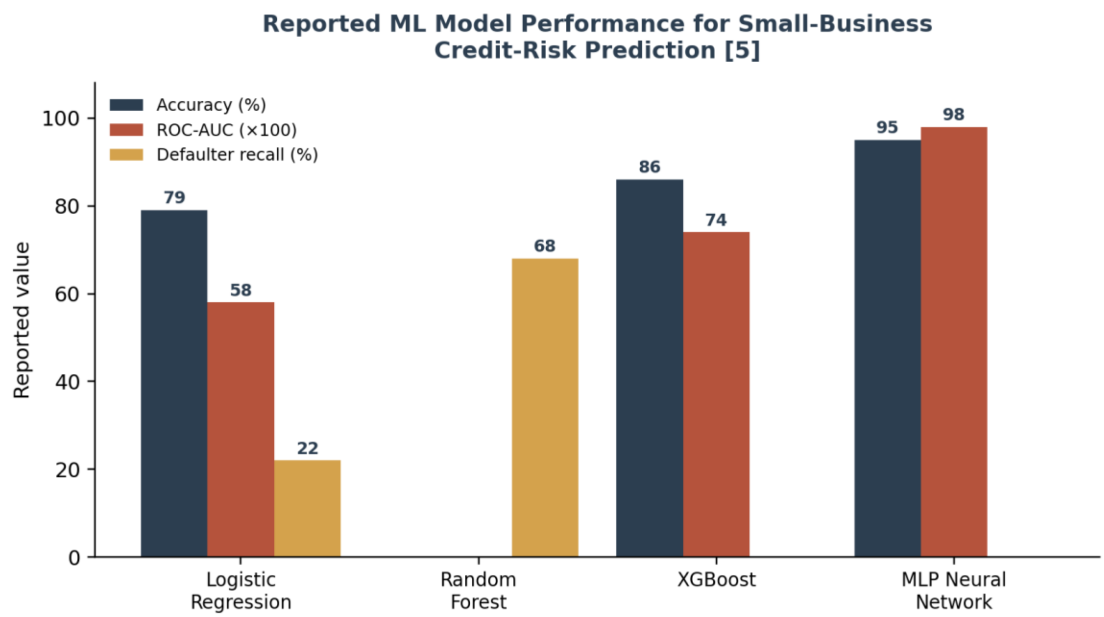


Figure 3 presents the comparative model results from the first study in Table 1 directly [5]. The pattern is notable and consistent with results reviewed elsewhere in this article series: neural-network and gradient-boosted ensemble methods substantially outperform logistic regression, the traditional workhorse of credit-scoring practice, on both overall accuracy and, critically for a distress-detection use case specifically, the ability to correctly identify actual defaulters (recall), where logistic regression's reported 22 percent defaulter recall would mean the model misses more than three-quarters of genuine defaults, a strikingly poor result for the specific early-warning purpose this article's architecture is designed to serve.

RELEVANCE OF EARLY-WARNING LITERATURE SPECIFICALLY

Beyond general credit-risk classification, a distinct strand of academic literature addresses early-warning systems for small-business distress specifically, predating the current machine-learning wave. Laitinen and Chong's foundational 1999 study, cited in subsequent interpretable-machine-learning research on SME credit default, established some of the earliest evidence on early-warning indicators for SME crisis using data from Finland and the UK, while Lin, Ansell, and Andreeva's 2012 study specifically examined how the definition of financial distress used to label training data affects small-business default-prediction results [7]. This latter finding is directly relevant to the architecture proposed in Section 3: how an institution defines the outcome its early-risk model is trained to predict, for example, 30-days delinquent versus a covenant breach versus a formal restructuring, materially affects both the model's behaviour and its practical usefulness, an institution-specific design choice rather than a fixed, universal parameter.

A WORKED ILLUSTRATIVE INTERPRETATION

To make Table 1's comparative results concrete for a CDFI credit officer, consider what the reported 22 percent defaulter recall for logistic regression in the comparative study [5] would mean in practice: of every ten small-business borrowers who will ultimately default, a logistic-regression-based screening model would be expected to correctly flag only about two in advance, leaving eight undetected until the default itself becomes apparent through missed payments. A model closer to the multi-layer-perceptron result in the same study, by contrast, would be expected to correctly flag a substantially higher share of eventual defaulters in advance, though the study does not report a directly comparable recall figure for that specific model. This gap illustrates concretely why the choice of underlying model matters for an early-warning use case specifically, as distinct from a simple credit-approval screening use case where overall accuracy alone might be a more acceptable proxy: an early-warning system that misses most genuine defaulters provides little practical early-warning value, regardless of its overall accuracy figure, since overall accuracy in a class-imbalanced default-prediction setting is dominated by the model's performance on the much larger population of borrowers who do not default.

DATA AND FEATURE CONSIDERATIONS FOR THE CDFI CONTEXT

The architecture in Figure 2 requires adaptation for the specific data constraints typical of the small-business-focused CDFI segment sized in Section 2. Given the small typical institution size documented in the 2023 CDFI Survey (23 percent of loan funds reporting 5 million U.S. dollars or less in total assets) [2], a typical small-business CDFI loan fund is unlikely to maintain the kind of large, in-house historical default dataset the studies in Table 1 were developed and validated against; the WSMOTE study's own emphasis on severe class-imbalance handling [6] is especially relevant here, since a small CDFI portfolio will typically contain very few confirmed default cases in absolute terms, exacerbating the class-imbalance challenge that broader-sample studies already identify as a central technical difficulty.

This data constraint has two direct architectural implications. First, a small CDFI loan fund may be better served initially by a simpler, more interpretable model, closer to the logistic-regression baseline in Table 1 despite its documented recall limitations, while it accumulates sufficient labelled outcome data to support a more sophisticated ensemble or neural-network approach, rather than attempting to deploy a data-hungry model against an insufficient training sample from the outset. Second, and consistent with the broader CDFI decision-support literature this article series addresses, transactional and operational data, available at higher frequency and lower marginal cost than accumulating additional years of confirmed default history, offer a more immediately actionable path to improving early-warning capability than waiting for a larger internal default dataset to accumulate through organic portfolio growth alone.

A FEATURE-ENGINEERING APPROACH SUITED TO THIN CDFI DATA

Given the class-imbalance and data-volume constraints documented above, feature engineering deserves particular emphasis for a CDFI-scale implementation, arguably more emphasis than model-architecture selection itself. Rather than relying primarily on the model to learn complex, non-linear patterns from a limited training sample, a CDFI-scale implementation should invest disproportionately in constructing well-reasoned, literature-informed features directly, drawing on the general indicator categories, liquidity trends, vendor-payment timeliness, credit-utilisation trends, that companion articles in this series detail more fully, and using the model itself primarily to combine these pre-engineered features rather than to discover them from raw data with limited examples to learn from. This approach trades some of the potential accuracy gain a data-rich model could achieve through automatic feature discovery for considerably more stable, interpretable behaviour on a small training sample, a trade-off well suited to the typical small-business CDFI's data constraints documented in Section 2.

DATA-SHARING POTENTIAL WITHIN THE CDFI SECTOR

A further, CDFI-specific consideration follows directly from the sector's own institutional structure: unlike conventional community banks, which generally operate as standalone competitors, CDFIs frequently participate in intermediary networks, industry associations, and shared-capital arrangements explicitly designed to support collective sector capacity, precisely the kind of existing institutional relationship that the federated and multi-institution validation architectures discussed elsewhere in this article series depend on. A CDFI-sector data-sharing initiative, pooling anonymised loan-performance data across willing participating loan funds specifically to build the shared training and validation dataset current individual institutions lack, would align naturally with this existing collaborative culture, and could plausibly progress more quickly than an equivalent effort among conventional, competing community banks, given CDFIs' shared mission orientation and pre-existing habit of sector-level collaboration through bodies such as Opportunity Finance Network.

EVIDENCE ASSESSMENT

Table 2 summarises the evidence strength for the principal claims this article relies on.

Table 2. Evidence-Strength Assessment for Key Claims in This Article

Claim	Evidence Status
CDFI industry scale and rapid five-year growth (1,487 institutions, $452B assets, tripled since 2018)	Well established — primary Federal Reserve Bank of New York research [1]
Sustained CDFI capacity strain despite growth (only 40% of CDFIs with rising demand fully meeting it)	Well established — primary Federal Reserve 2023 CDFI Survey data [2]
Sustained/growing CDFI small-business demand (75% reporting increased demand)	Well established — primary Federal Reserve 2023 CDFI Survey data [2]
MLP neural networks achieve 95% accuracy / 0.98 AUC for small-business credit risk	Reported in industry-analyst source [5]; direction consistent with, but this specific figure not independently cross-validated against, peer-reviewed academic literature in this article's research
WSMOTE ensemble improves minority-class accuracy by 15.16%	Peer-reviewed, published result, though evaluated on general small-business loan data rather than CDFI-specific data [6]
The proposed architecture, or any of the reviewed models, has been validated on CDFI-originated loan data specifically	Not evidenced in any source identified for this article — this is the central, unaddressed gap this article's evidence review surfaces

The single most important limitation to state plainly is the one summarised in the table's final row: no study identified in this article's research evaluates a machine-learning early-risk model against CDFI-originated small-business loan data specifically. This matters because CDFI borrowers are, by the sector's own mission and the evidence reviewed in Section 2, systematically different from the broader small-business borrower populations underlying the studies in Table 1, generally thinner credit files, less collateral, and greater reliance on qualitative relationship knowledge in underwriting, meaning a model's published performance on a general small-business dataset should not be assumed to transfer directly to a CDFI's specific borrower population without independent, CDFI-specific validation.

GOVERNANCE AND ADOPTION CONSIDERATIONS

Given the scale constraints documented in Section 2 and the evidence gap documented in Section 6, a phased, cautious adoption path is appropriate for most small-business-focused CDFIs. A reasonable starting point, consistent with the data-constraint discussion in Section 5, is a rules-informed scorecard using interpretable features and published, literature-informed indicator weightings as an interim measure, backtested retrospectively against whatever historical portfolio outcomes the institution already has available, before progressing to a more sophisticated ensemble or neural-network model once sufficient CDFI-specific outcome data has accumulated to support independent validation. Given the specific fair-lending sensitivity inherent to CDFIs' mission of serving historically underserved borrowers, any model adopted, at any stage of this phased path, should be tested explicitly for disparate impact across protected classes and geographies before deployment and on an ongoing basis thereafter, consistent with the governance principles discussed in companion articles in this series.

Given that many small-business CDFIs will lack the internal capacity to conduct this kind of validation independently, industry associations and CDFI intermediaries, of the kind referenced in the multi-institution validation literature this article series addresses elsewhere, are a natural candidate to sponsor a shared, CDFI-specific validation effort, pooling anonymised outcome data across multiple small-business-focused CDFIs to build exactly the kind of CDFI-specific evidence base Section 6 identifies as currently missing, an effort that individual small CDFI loan funds, given the scale documented in Section 2, are unlikely to be able to undertake independently.

BOARD AND FUNDER COMMUNICATION

Beyond internal governance, small-business CDFIs face a distinctive external reporting obligation that conventional community banks generally do not: many CDFIs rely substantially on philanthropic and public grant funding, per the funding-source discussion in Section 2.1, and their funders increasingly expect evidence of data-driven, efficient capital deployment as a condition of continued support. A well-documented early-risk-signaling implementation, with the kind of governance architecture proposed in Section 4 and the honest evidence-strength assessment modelled in Section 6, can serve a dual purpose for a small-business CDFI specifically: strengthening actual credit-risk management while also providing funders with concrete, defensible evidence of the institution's operational sophistication and stewardship of scarce capital, a consideration with particular salience given that many funders now expect data-driven evidence of efficient capital deployment as the sector continues to expand [2].

CONCLUSION

Before summarising, it is worth restating the specific decision this article's evidence review is best positioned to inform: whether a small-business-focused CDFI should begin investing in early-risk-signaling analytics at all, given its resource constraints, and if so, along what sequence. The evidence in Table 1 supports beginning, since the underlying technique class is well validated, even if not yet CDFI-specifically, and the evidence in Section 2 supports urgency, since the sector's rapid growth and sustained small-business demand together suggest that the value of extending scarce CDFI staff capacity through analytics is more likely to grow than shrink over the coming years, not less. The sequencing question, addressed throughout Sections 3 through 7, points toward a cautious, phased path beginning with interpretable, literature-informed scoring and existing rule-based screening, rather than an immediate, full-scale deployment of the most sophisticated model reviewed in Section 4, given the CDFI-specific validation gap Section 6 documents plainly. This article has proposed a decision-support architecture for embedding early-risk-signaling analytics into CDFI credit-risk and loan-monitoring workflows, grounded in published, cited evidence on machine-learning approaches to small-business credit-risk prediction and in current, real data on the CDFI industry's scale, rapid recent growth, and expanding role in small-business lending. The evidence reviewed establishes clearly that machine-learning methods, particularly neural-network and gradient-boosted ensemble approaches, substantially outperform traditional logistic-regression credit scoring on both overall accuracy and, more importantly for an early-warning use case, the ability to correctly identify genuine defaulters [5][6][8]. It establishes equally clearly, however, that this evidence derives entirely from general small-business and SME credit data, not from CDFI-originated loans specifically, leaving CDFI-specific validation as a genuine, currently unaddressed gap this article has been explicit about throughout rather than glossing over. Given the sector-wide growth and capacity pressure documented in Section 2, a shared, CDFI-sector validation effort, of the kind proposed in Section 7 and consistent with the multi-institution validation approach this article series recommends elsewhere, represents a particularly time-sensitive priority: a rapidly growing, still resource-constrained CDFI industry serving sustained small-business demand is precisely the context in which analytics-driven efficiency gains offer the most value, and precisely the context in which individual institutions are least likely to be able to generate the necessary validating evidence independently.

REFERENCES

[1] Bazzana, F., Bee, M., & Khatir, A. A. H. A. (2024). Machine learning techniques for default prediction: An application to small Italian companies. *Risk Management, 26*, Article 1.

[2] Bitetto, A., Cerchiello, P., Filomeni, S., Tanda, A., & Tarantino, B. (2024). Can we trust machine learning to predict the credit risk of small businesses? *Review of Quantitative Finance and Accounting, 63*, 925–954.

[3] Dendramis, Y., Tzavalis, E., & Cheimarioti, A. (2023). Measuring the default risk of small business loans: Improved credit risk prediction using deep learning. *Journal of Forecasting, 44*(7), 2277–2297.

[4] Dimitrova-Grajzl, V., Grajzl, P., Guse, A. J., & Kokodoko, M. (2022). Community development financial institutions and individuals' credit risk in Indian Country. *Journal of Economic Issues, 56*(1), 184–208.

[5] Dimitrova-Grajzl, V., Grajzl, P., Guse, J., Kokodoko, M., & Wheeler, L. (2024). Beyond conventional models: Lending by Native Community Development Financial Institutions. *Annals of Public and Cooperative Economics, 95*, 675–701.

[6] Huang, B., Zhao, F., Tian, M., Zhang, D., Zhang, X., Wang, Z., Li, H., & Chen, B. (2024). Explainability of machine learning in credit risk assessment of SMEs. *Frontiers in Artificial Intelligence and Applications*.

[7] Lehn, B., Rubin, J. S., & Zielenbach, S. (2004). Community development financial institutions: Current issues and future prospects. *Journal of Urban Affairs, 26*(2), 177–195.

[8] Stevenson, M., Mues, C., & Bravo, C. (2021). The value of text for small business default prediction: A deep learning approach. *European Journal of Operational Research, 295*(2), 758–771.